GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2005, 07:40:07 ; Search time 9801 Seconds
(without alignments)
4454.459 Million cell updates/sec

Title: US-09-920-705-3
Perfect score: 4801
Sequence: 1 MALSMKPFRADDSGFQSNNL.....LQIQQKQQQQSSGTPAGQGP 901

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09920705/runat_07042005_101813_20638/app_query.fasta_1.1095
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09920705_@CGN_1_1_5932_@runat_07042005_101813_20638 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4801	100.0	2706	6	AX420420	AX420420 Sequence
2	4786	99.7	2706	8	BT003029	BT003029 Arabidops
3	4786	99.7	3498	8	AY094463	AY094463 Arabidops
4	4656.5	97.0	3171	6	CQ804588	CQ804588 Sequence
5	4392	91.5	4593	6	AX420419	AX420419 Sequence
c 6	4392	91.5	102057	8	AC002335	AC002335 Arabidops
c 7	1436	29.9	801	6	AX420422	AX420422 Sequence
8	1365	28.4	3543	8	AK065222	AK065222 Oryza sat
c 9	1117.5	23.3	124083	8	AP005735	AP005735 Oryza sat
c 10	1117.5	23.3	154228	8	AC108755	AC108755 Oryza sat
c 11	1117.5	23.3	176553	8	AP006756	AP006756 Oryza sat
12	799	16.6	3455	8	AK100889	AK100889 Oryza sat
13	651.5	13.6	3715	6	AX420424	AX420424 Sequence
c 14	576	12.0	130961	8	AP004011	AP004011 Oryza sat
15	573.5	11.9	137519	8	ATT1P17	AL049730 Arabidops
16	573.5	11.9	196107	8	ATCHRIV34	AL161534 Arabidops
17	466	9.7	349	6	AX420423	AX420423 Sequence
c 18	374.5	7.8	116419	8	AC146743	AC146743 Medicago
19	332.5	6.9	6836	9	HSA303089	AJ303089 Homo sapi
20	331.5	6.9	6770	9	HSA303090	AJ303090 Homo sapi
21	302	6.3	2190	6	BD223778	BD223778 Human RNA
22	302	6.3	6057	9	HSM803154	AL831838 Homo sapi
23	302	6.3	168499	9	AC092037	AC092037 Homo sapi
c 24	302	6.3	174525	2	AC067763	AC067763 Homo sapi
25	297	6.2	1692	6	CQ736847	CQ736847 Sequence
26	287.5	6.0	2308	3	AY069434	AY069434 Drosophil
27	287.5	6.0	3084	6	CQ577575	CQ577575 Sequence
28	287.5	6.0	3113	3	AF145664	AF145664 Drosophil
c 29	287.5	6.0	9744	6	CQ577574	CQ577574 Sequence
30	287.5	6.0	37983	2	AC020394	AC020394 Drosophil
31	287.5	6.0	86889	3	AC005448	AC005448 Drosophil
c 32	287.5	6.0	139410	3	AC007084	AC007084 Drosophil
33	287.5	6.0	254023	3	AE003839	AE003839 Drosophil
34	277	5.8	10839	10	AF156529	AF156529 Mus muscu
35	276.5	5.8	3384	9	AF368064	AF368064 Homo sapi
36	273	5.7	3192	6	BD159584	BD159584 Primer fo
37	273	5.7	3192	6	AX882187	AX882187 Sequence
38	273	5.7	3192	9	AK022541	AK022541 Homo sapi
39	273	5.7	3312	9	AF368062	AF368062 Homo sapi
40	273	5.7	3423	9	AF368063	AF368063 Homo sapi
41	273	5.7	4192	5	BC070824	BC070824 Xenopus l
42	271.5	5.7	7815	9	HSA297259	AJ297259 Homo sapi
43	271.5	5.7	10718	10	AB055980	AB055980 Mus muscu
44	271.5	5.7	81137	9	AL355488	AL355488 Human DNA
c 45	271.5	5.7	158483	2	AC025987	AC025987 Homo sapi

ALIGNMENTS

RESULT 1
AX420420
LOCUS AX420420 2706 bp DNA linear PAT 18-JUN-2002
DEFINITION Sequence 2 from Patent WO0212518.
ACCESSION AX420420
VERSION AX420420.1 GI:21524576
KEYWORDS .
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Amasino,R.M., Schomburg,F.M., Michaels,S.D. and Patton,D.
TITLE Floral induction gene
JOURNAL Patent: WO 0212518-A 2 14-FEB-2002;
WISCONSIN ALUMNI RESEARCH FOUNDATION (US)
FEATURES Location/Qualifiers
source 1. .2706
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
CDS 1. .2706
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD35485.1"
/db_xref="GI:21524577"
/translation="MALSMKPFRADDSGFQSNNLWVGSLTPETTESDLTELFGRYGDI

```
                     DRITVYSSRGFAFIYYRHVEEAVAAKEALQGANLNGSQIKIEYARPAKPCKSLWVGGI
                     GPNVSKDDLEEEFSKFGKIEDFRFLRERKTAFIDYYEMDDALQAKSMNGKPMGGSFLR
                     VDFLRSQAPKKEQWAGSYDNRNGNMNHKPQYPHSYEDFKGDVQPSKVLWIGFPPTATQ
                     CNDEQILHNAMILFGEIERVKSYPSRNFALVEFRSAEEARQCKEGLQGRLFNNPRIKI
                     MYSNDELPPEQDDTSFYSGMKRSRTDMFNNDPSCVSSPHSTGIPGSMRPLRGTNERSY
                     NGAEYNDVVGKEPNWRRPSANGTGILPSPTGPGILPSPAQGTRRPMRSNPDSWEGYDP
                     AQLVRESKRTRRDGSVDGFTPMGVDERSFGRGSVAARPIRGPPDSDHIWRGMIAKGGT
                     PVCCARCVPMGKGIETKLPEVVNCSARTDLNMLAKHYAVAIGCEIVFFVPDREEDFAS
                     YTEFLRYLSSKDRAGVAKLDDGTTLFLVPPSDFLTDVLQVTRQERLYGVVLKLPPPAV
                     PVTASYRQESQSNPLHYMDQARDSPANASHSLYPPRENYIRGAPEHLTAASKPSVSEP
                     LRIPNNAAPQAGVSLTPELLATLASILPATSQPAAPESHQPMSGPSTVVSTAHQSNGL
                     YNGEAPSQAWKRGPQTVHDASNQSFQQYGNQYTPAGQLPPPPSRYPPASNNPNYTSGM
                     VHGNMQYQSQSVNMPQLSPLPNMPHNNYSMYTQGSSNHPVSQPMVQQYQPEASMPNQN
                     YGPIPSYQQANFHGVTTNQAQNLNPSQFQAAMQPPADKANLEPQNQALRLQPMISGDG
                     QGTTDGEVDKNQRYQSTLQFAANLLLQIQQKQQQQSSGTPAGQGP"
     misc_binding    46. .279
                     /bound_moiety="RNA Binding Region"
     misc_binding    283. .522
                     /bound_moiety="RNA Binding Region"
     misc_binding    610. .852
                     /bound_moiety="RNA Binding Region"
ORIGIN

Alignment Scores:
Pred. No.:                9.13e-184      Length:          2706
Score:                    4801.00        Matches:         901
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       6              Gaps:            0

US-09-920-705-3 (1-901) x AX420420 (1-2706)

Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 60

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 120

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 180

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 240

Qy         81 SerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTCAAATTAAGATCGAATACGCACGACCGGCAAAACCTTGTAAGAGTCTATGGGTGGGT 300

Qy        101 GlyIleGlyProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAATCGGCCCTAATGTCTCCAAGGATGACCTGGAGGAAGAGTTCAGCAAGTTTGGGAAA 360

Qy        121 IleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMet 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATCGAGGATTTTAGGTTTCTCAGAGAACGCAAGACAGCTTTCATTGATTATTATGAGATG 420

Qy        141 AspAspAlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArg 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATGATGCTTTACAGGCTAAGAGCATGAATGGAAAGCCTATGGGTGGTAGCTTTTTGCGT 480

Qy        161 ValAspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsn 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTTGATTTTCTCCGGTCACAAGCGCCAAAAAAAGAACAATGGGCTGGCTCTTACGATAAC 540

Qy        181 ArgAsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGly 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGAAATGGCAATATGAATCATAAACCGCAGTATCCTCACTCATATGAAGACTTTAAAGGA 600

Qy        201 AspValGlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GATGTCCAGCCAAGTAAGGTTCTGTGGATTGGGTTCCCTCCTACTGCTACACAATGCAAT 660

Qy        221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATGAGCAAATTCTGCACAATGCGATGATACTCTTTGGTGAGATCGAGAGGGTAAAAAGT 720

Qy        241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TACCCATCAAGGAATTTTGCACTTGTGGAGTTTAGGAGCGCGGAGGAAGCTCGCCAATGC 780

Qy        261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGGAAGGCCTACAGGGGAGGTTATTCAATAATCCTAGAATCAAAATTATGTACTCAAAC 840

Qy        281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GATGAGTTGCCTCCTGAGCAAGACGATACTAGTTTTTACTCTGGTATGAAACGGTCAAGG 900

Qy        301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACAGATATGTTCAATAATGATCCTTCATGTGTATCTTCTCCTCATTCTACTGGAATTCCT 960

Qy        321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsn 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGGTCTATGAGGCCCCTCAGAGGTACGAATGAGCGTTCATATAATGGTGCAGAATACAAT 1020

Qy        341 AspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GACGTTGTTGGTAAGGAGCCAAACTGGAGGAGGCCATCTGCAAATGGAACTGGAATACTC 1080

Qy        361 ProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMet 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCATCTCCAACAGGACCTGGAATCCTCCCATCTCCTGCACAAGGTACGAGGCGCCCTATG 1140

Qy        381 ArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuValArgGluSerLys 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AGGTCAAACCCCGATTCTTGGGAAGGATATGATCCTGCTCAGTTGGTCAGAGAAAGTAAA 1200

Qy        401 ArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGluArgSer 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGAACCAGAAGAGATGGATCAGTGGACGGTTTTACTCCAATGGGTGTCGATGAGAGGTCA 1260

Qy        421 PheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAspSerAspHisIle 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTTGGTCGAGGTTCAGTTGCTGCTAGACCTATCCGTGGCCCCCCTGATTCTGATCACATA 1320

Qy        441 TrpArgGlyMetIleAlaLysGlyGlyThrProValCysCysAlaArgCysValProMet 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGAGAGGAATGATTGCCAAGGGTGGAACTCCCGTCTGTTGTGCTCGTTGTGTACCTATG 1380

Qy        461 GlyLysGlyIleGluThrLysLeuProGluValValAsnCysSerAlaArgThrAspLeu 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GGAAAGGGGATTGAAACTAAACTGCCTGAGGTCGTCAATTGTTCAGCAAGAACTGATTTG 1440

Qy        481 AsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGluIleValPhePheValPro 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AATATGCTCGCTAAACATTACGCCGTTGCCATTGGATGTGAGATCGTTTTTTTCGTACCA 1500

Qy        501 AspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAsp 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GACAGGGAAGAAGATTTTGCGTCTTACACTGAATTTCTCCGGTACCTTAGCTCAAAAGAT 1560

Qy        521 ArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProProSerAsp 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CGGGCGGGTGTTGCCAAATTAGATGATGGTACAACTTTATTCTTGGTGCCTCCATCAGAT 1620

Qy        541 PheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLys 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTTAACTGATGTACTCCAAGTGACCCGTCAAGAACGGCTATATGGTGTTGTTCTCAAG 1680

Qy        561 LeuProProProAlaValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnPro 580
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTACCCCCGCCAGCCGTTCCTGTTACAGCATCATACAGACAAGAATCTCAGTCCAATCCT 1740

Qy        581 LeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrPro 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTGCATTATATGGATCAAGCCCGGGATTCACCTGCCAATGCTAGTCACAGTTTATATCCT 1800

Qy        601 ProArgGluAsnTyrIleArgGlyAlaProGluHisLeuThrAlaAlaSerLysProSer 620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCTAGGGAAAATTACATTAGGGGTGCACCAGAACATTTGACAGCTGCTTCAAAACCATCT 1860

Qy        621 ValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThr 640
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GTTAGCGAGCCTCTCAGAATACCTAATAATGCAGCGCCTCAAGCTGGGGTTAGTTTAACT 1920

Qy        641 ProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCGGAGCTTTTAGCCACTCTGGCATCTATTCTCCCTGCAACTTCTCAACCTGCTGCCCCT 1980

Qy        661 GluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsn 680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GAGAGTCACCAACCTATGTCAGGACCTTCAACAGTTGTTTCCACAGCACATCAGTCCAAT 2040

Qy        681 GlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHis 700
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GGACTGTACAATGGAGAAGCACCGTCTCAAGCTTGGAAAAGAGGTCCACAAACAGTTCAT 2100

Qy        701 AspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeu 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GATGCGTCAAATCAGTCATTCCAACAATACGGAAATCAGTACACTCCAGCTGGGCAACTA 2160

Qy        721 ProProProProSerArgTyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMet 740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CCTCCTCCTCCTTCGCGTTACCCTCCAGCTTCAAACAACCCCAACTACACTAGTGGAATG 2220

Qy        741 ValHisGlyAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeu 760
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GTCCATGGCAACATGCAATACCAGAGCCAATCTGTTAACATGCCTCAGCTGTCTCCGTTA 2280

Qy        761 ProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProVal 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCAAATATGCCTCATAATAATTATTCCATGTACACTCAGGGTTCGTCAAATCATCCTGTT 2340

Qy        781 SerGlnProMetValGlnGlnTyrGlnProGluAlaSerMetProAsnGlnAsnTyrGly 800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TCTCAGCCCATGGTCCAGCAATACCAACCAGAAGCGTCCATGCCAAACCAAAACTATGGT 2400

Qy        801 ProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsn 820
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 CCAATTCCAAGTTATCAGCAAGCTAATTTTCATGGCGTAACAACAAATCAGGCACAGAAC 2460

Qy        821 LeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGlu 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 TTAAACCCTTCCCAATTTCAAGCTGCCATGCAACCACCAGCAGATAAGGCAAATTTAGAG 2520

Qy        841 ProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThr 860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 CCACAAAACCAAGCACTACGATTGCAGCCTATGATCTCTGGGGATGGTCAGGGTACAACA 2580

Qy        861 AspGlyGluValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsnLeu 880
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 GATGGGGAGGTCGATAAGAATCAGAGATACCAGTCAACACTACAATTTGCAGCAAACCTT 2640

Qy        881 LeuLeuGlnIleGlnGlnLysGlnGlnGlnGlnSerSerGlyThrProAlaGlyGlnGly 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CTTCTCCAGATACAGCAGAAACAGCAGCAACAGTCTTCAGGTACTCCGGCTGGACAGGGG 2700

Qy        901 Pro 901
                |||
Db       2701 CCT 2703

RESULT 2
BT003029
```

```
LOCUS       BT003029               2706 bp    mRNA    linear   PLN 16-JAN-2003
DEFINITION  Arabidopsis thaliana At2g43410/T1O24.15 gene, complete cds.
ACCESSION   BT003029
VERSION     BT003029.1  GI:27764945
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2706)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
            Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K.,
            Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and
            Ecker,J.R.
  TITLE     Arabidopsis ORF clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2706)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
            Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K.,
            Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and
            Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2003) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') :  Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P.,  Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members constructed and
            sequenced the pUNI (ORF) clones using the RAFL cDNAs: Cheuk,R.,
            Chen,H., Kim,C.J., Shinn,P., Bowser,L., Chan, M.M., Chang,C.M.,
            Dale,J.M., Hsuan,V.W., Jones,T., Karlin-Neumann,G., Lam,B.,
            Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera, C.S., Palm,C.J.,
            Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
            Yamada,K., Yu,G., Yuan,S., Davis,R.W.,  Theologis,A., and
            Ecker,J.R.

            Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .2706
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /clone="U21966"
                     /ecotype="Columbia"
                     /note="This clone is in pUNI 51"
     CDS             1. .2706
                     /note="unknown protein"
                     /codon_start=1
                     /product="At2g43410/T1O24.15"
                     /protein_id="AAO23594.1"
                     /db_xref="GI:27764946"
                     /translation="MALSMKPFRADDSGFQSNNLWVGSLTPETTESDLTELFGRYGDI
                     DRITVYSSRGFAFIYYRHVEEAVAAKEALQGANLNGSQIKIEYARPAKPCKSLWVGGI
                     GPNVSKDDLEEEFSKFGKIEDFRFLRERKTAFIDYYEMDDALQAKSMNGKPMGGSFLR
                     VDFLRSQAPKKEQWAGSYDNRNGNMNHKPQYPHSYEDFKGDVQPSKVLWIGPPPTATQ
                     CNDEQILHNAMILFGEIERVKSYPSRNFALVEFRSAEEARQCKEGLQGRLFNNPRIKI
```

MYSNDELPPEQDDTSFYSGMKRSRTDMFNNDPSFVSSPHSTGIPGSMRPLRGTNERSY
NGAEYNDVVGKEPNWRRPSANGTGILPSPTGPGILPSPAQGTRRPMRSNPDSWEGYDP
AQLVRESKRTRRDGSVDGFTPMGVDERSFGRGSVAARPIRGPPDSDHIWRGMIAKGGT
PVCCARCVPMGKGIETKLPEVVNCSARTDLNMLAKHYAVAIGCEIVFFVPDREEDFAS
YTEFLRYLSSKDRAGVAKLDDGTTLFLVPPSDFLTDVLQVTRQERLYGVVLKLPPPAV
PVTASYRQESQSNPLHYMDQARDSPANASHSLYPPRENYIRGAPEHLTAASKPSVSEP
LRIPNNAAPQAGVSLTPELLATLASILPATSQPAAPESHQPMSGPSTVVSTAHQSNGL
YNGEAPSQAWKRGPQTVHDASNQSFQQYGNQYTPAGQLPPPPSRYPPASNNPNYTSGM
VHGNMQYQSQSVNMPQLSPLPNMPHNNYSMYTQGSSNHPVSQPMVQQYQPEASMPNQN
YGPIPSYQQANFHGVTTNQAQNLNPSQFQAAMQPPADKANLEPQNQALRLQPMISGDG
QGTTDGEVDKNQRYQSTLQFAANLLLQIQQKRQQQSSGTPAGQGP"

ORIGIN

Alignment Scores:

Pred. No.:	3.64e-183	Length:	2706
Score:	4786.00	Matches:	899
Percent Similarity:	99.89%	Conservative:	1
Best Local Similarity:	99.78%	Mismatches:	1
Query Match:	99.69%	Indels:	0
DB:	8	Gaps:	0

US-09-920-705-3 (1-901) x BT003029 (1-2706)

```
Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu  20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT  60

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg  40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 120

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 180

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly  80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 240

Qy         81 SerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTCAAATTAAGATCGAATACGCACGACCGGCAAAACCTTGTAAGAGTCTATGGGTGGGT 300

Qy        101 GlyIleGlyProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAATCGGCCCTAATGTCTCCAAGGATGACCTGGAGGAAGAGTTCAGCAAGTTTGGGAAA 360

Qy        121 IleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMet 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATCGAGGATTTTAGGTTTCTCAGAGAACGCAAGACAGCTTTCATTGATTATTATGAGATG 420

Qy        141 AspAspAlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArg 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATGATGCTTTACAGGCTAAGAGCATGAATGGAAAGCCTATGGGTGGTAGCTTTTTGCGT 480

Qy        161 ValAspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsn 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTTGATTTTCTCCGGTCACAAGCGCCAAAAAAAGAACAATGGGCTGGCTCTTACGATAAC 540

Qy        181 ArgAsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGly 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGAAATGGCAATATGAATCATAAACCGCAGTATCCTCACTCATATGAAGACTTTAAAGGA 600

Qy        201 AspValGlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GATGTCCAGCCAAGTAAGGTTCTGTGGATTGGGTTCCCTCCTACTGCTACACAATGCAAT 660

Qy        221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATGAGCAAATTCTGCACAATGCGATGATACTCTTTGGTGAGATCGAGAGGGTAAAAAGT 720

Qy        241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        721 TACCCATCAAGGAATTTTGCACTTGTGGAGTTTAGGAGCGCGGAGGAAGCTCGCCAATGC 780

Qy        261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGGAAGGCCTACAGGGGAGGTTATTCAATAATCCTAGAATCAAAATTATGTACTCAAAC 840

Qy        281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GATGAGTTGCCTCCTGAGCAAGACGATACTAGTTTTTACTCTGGTATGAAACGGTCAAGG 900

Qy        301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACAGATATGTTCAATAATGATCCTTCATTTGTATCTTCTCCTCATTCTACTGGAATTCCT 960

Qy        321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsn 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGGTCTATGAGGCCCCTCAGAGGTACAAATGAGCGTTCATATAATGGTGCAGAATACAAT 1020

Qy        341 AspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GACGTTGTTGGTAAGGAGCCAAACTGGAGGAGGCCATCTGCAAATGGAACTGGAATACTC 1080

Qy        361 ProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMet 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCATCTCCAACAGGACCTGGAATCCTCCCATCTCCTGCACAAGGTACGAGGCGCCCTATG 1140

Qy        381 ArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuValArgGluSerLys 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AGGTCAAACCCCGATTCTTGGGAAGGATATGATCCTGCTCAGTTGGTCAGAGAAAGTAAA 1200

Qy        401 ArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGluArgSer 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGAACCAGAAGAGATGGATCAGTGGACGGTTTTACTCCAATGGGTGTCGATGAGAGGTCA 1260

Qy        421 PheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAspSerAspHisIle 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTTGGTCGAGGTTCAGTTGCTGCTAGACCTATCCGTGGCCCCCCTGATTCTGATCACATA 1320

Qy        441 TrpArgGlyMetIleAlaLysGlyGlyThrProValCysCysAlaArgCysValProMet 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGAGAGGAATGATTGCCAAGGGTGGAACTCCCGTCTGTTGTGCTCGTTGTGTACCTATG 1380

Qy        461 GlyLysGlyIleGluThrLysLeuProGluValValAsnCysSerAlaArgThrAspLeu 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GGAAAGGGGATTGAAACTAAACTGCCTGAGGTCGTCAATTGTTCAGCAAGAACTGATTTG 1440

Qy        481 AsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGluIleValPhePheValPro 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AATATGCTCGCTAAACATTACGCCGTTGCCATTGGATGTGAGATCGTTTTTTTCGTACCA 1500

Qy        501 AspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAsp 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GACAGGGAAGAAGATTTTGCGTCTTACACTGAATTTCTCCGGTACCTTAGCTCAAAAGAT 1560

Qy        521 ArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProProSerAsp 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CGGGCGGGTGTTGCCAAATTAGATGATGGTACAACTTTATTCTTGGTGCCTCCATCAGAT 1620

Qy        541 PheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLys 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTTAACTGATGTACTCCAAGTGACCCGTCAAGAACGGCTATATGGTGTTGTTCTCAAG 1680

Qy        561 LeuProProProAlaValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnPro 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTACCCCCGCCAGCCGTTCCTGTTACAGCATCATACAGACAAGAATCTCAGTCCAATCCT 1740

Qy        581 LeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrPro 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTGCATTATATGGATCAAGCCCGGGATTCACCTGCCAATGCTAGTCACAGTTTATATCCT 1800

Qy        601 ProArgGluAsnTyrIleArgGlyAlaProGluHisLeuThrAlaAlaSerLysProSer 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCTAGGGAAAATTACATTAGGGGTGCACCAGAACATTTGACAGCTGCTTCAAAACCATCT 1860
```

```
Qy        621 ValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThr 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GTTAGCGAGCCTCTCAGAATACCTAATAATGCAGCGCCTCAAGCTGGGGTTAGTTTAACT 1920

Qy        641 ProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCGGAGCTTTTAGCCACTCTGGCATCTATTCTCCCTGCAACTTCTCAACCTGCTGCCCCT 1980

Qy        661 GluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsn 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GAGAGTCACCAACCTATGTCAGGACCTTCAACAGTTGTTTCCACAGCACATCAGTCCAAT 2040

Qy        681 GlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHis 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GGACTGTACAATGGAGAAGCACCGTCTCAAGCTTGGAAAAGAGGTCCACAAACAGTTCAT 2100

Qy        701 AspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeu 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GATGCGTCAAATCAGTCATTCCAACAATACGGAAATCAGTACACTCCAGCTGGGCAACTA 2160

Qy        721 ProProProProSerArgTyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMet 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CCTCCTCCTCCTTCGCGTTACCCTCCAGCTTCAAACAACCCCAACTACACTAGTGGAATG 2220

Qy        741 ValHisGlyAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeu 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GTCCATGGCAACATGCAATACCAGAGCCAATCTGTTAACATGCCTCAGCTGTCTCCGTTA 2280

Qy        761 ProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProVal 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCAAATATGCCTCATAATAATTATTCCATGTACACTCAGGGTTCGTCAAATCATCCTGTT 2340

Qy        781 SerGlnProMetValGlnGlnTyrGlnProGluAlaSerMetProAsnGlnAsnTyrGly 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TCTCAGCCCATGGTCCAGCAATACCAACCAGAAGCGTCCATGCCAAACCAAAACTATGGT 2400

Qy        801 ProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsn 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 CCAATTCCAAGTTATCAGCAAGCTAATTTTCATGGCGTAACAACAAATCAGGCACAGAAC 2460

Qy        821 LeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGlu 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 TTAAACCCTTCCCAATTTCAAGCTGCCATGCAACCACCAGCAGATAAGGCAAATTTAGAG 2520

Qy        841 ProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThr 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 CCACAAAACCAAGCACTACGATTGCAGCCTATGATCTCTGGGGATGGTCAGGGTACAACA 2580

Qy        861 AspGlyGluValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsnLeu 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 GATGGGGAGGTCGATAAGAATCAGAGATACCAGTCAACACTACAATTTGCAGCAAACCTT 2640

Qy        881 LeuLeuGlnIleGlnGlnLysGlnGlnGlnGlnSerSerGlyThrProAlaGlyGlnGly 900
              ||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db       2641 CTTCTCCAGATACAGCAGAAACGGCAGCAACAGTCTTCAGGTACTCCGGCTGGACAGGGG 2700

Qy        901 Pro 901
              |||
Db       2701 CCT 2703

RESULT 3
AY094463
LOCUS       AY094463               3498 bp    mRNA    linear   PLN 05-MAY-2002
DEFINITION  Arabidopsis thaliana At2g43410/T1O24.15 mRNA, complete cds.
ACCESSION   AY094463
VERSION     AY094463.1  GI:20453184
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE   1  (bases 1 to 3498)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Meyers,M.C., Banh,J.,
            Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D.,
            Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M.,
            Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T.,
            Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W.,
            Theologis,A. and Ecker,J.R.
  TITLE     Arabidopsis cDNA clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3498)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Meyers,M.C., Banh,J.,
            Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D.,
            Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M.,
            Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T.,
            Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W.,
            Theologis,A. and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') :  Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P.,  Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Shinn,P., Chen,H.,
            Cheuk,R., Kim,C.J.,  Meyers,M.C., Banh,J. Bowser,L., Chang,E.,
            Dale,J.M., Goldsmith,A.D., Jones,T., Karlin-Neumann,G., Lam,B.,
            Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J.,
            Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamada,K., Yamamura, Y., Yu,G., Yu,S., Davis,R.W.,  Theologis,A.,
            and Ecker,J.R.

            Shinn,P. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .3498
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /clone="RAFL09-47-P10(R21966)"
                     /ecotype="Columbia"
     CDS             255. .2960
                     /note="unknown protein"
                     /codon_start=1
                     /product="At2g43410/T1O24.15"
                     /protein_id="AAM19833.1"
                     /db_xref="GI:20453185"
                     /translation="MALSMKPFRADDSGFQSNNLWVGSLTPETTESDLTELFGRYGDI
                     DRITVYSSRGFAFIYYRHVEEAVAAKEALQGANLNGSQIKIEYARPAKPCKSLWVGGI
                     GPNVSKDDLEEEFSKFGKIEDFRFLRERKTAFIDYYEMDDALQAKSMNGKPMGGSFLR
                     VDFLRSQAPKKEQWAGSYDNRNGNMNHKPQYPHSYEDFKGDVQPSKVLWIGFPPTATQ
                     CNDEQILHNAMILFGEIERVKSYPSRNFALVEFRSAEEARQCKEGLQGRLFNNPRIKI
                     MYSNDELPPEQDDTSFYSGMKRSRTDMFNNDPSFVSSPHSTGIPGSMRPLRGTNERSY
                     NGAEYNDVVGKEPNWRRPSANGTGILPSPTGPGILPSPAQGTRRPMRSNPDSWEGYDP
                     AQLVRESKRTRRDGSVDGFTPMGVDERSFGRGSVAARPIRGPPDSDHIWRGMIAKGGT
                     PVCCARCVPMGKGIETKLPEVVNCSARTDLNMLAKHYAVAIGCEIVFFVPDREEDFAS
                     YTEFLRYLSSKDRAGVAKLDDGTTLFLVPPSDFLTDVLQVTRQERLYGVVLKLPPPAV
                     PVTASYRQESQSNPLHYMDQARDSPANASHSLYPPRENYIRGAPEHLTAASKPSVSEP
                     LRIPNNAAPQAGVSLTPELLATLASILPATSQPAAPESHQPMSGPSTVVSTAHQSNGL
                     YNGEAPSQAWKRGPQTVHDASNQSFQQYGNQYTPAGQLPPPPSRYPPASNNPNYTSGM
                     VHGNMQYQSQSVNMPQLSPLPNMPHNNYSMYTQGSSNHPVSQPMVQQYQPEASMPNQN
                     YGPIPSYQQANFHGVTTNQAQNLNPSQFQAAMQPPADKANLEPQNQALRLQPMISGDG
                     QGTTDGEVDKNQRYQSTLQFAANLLLQIQQKRQQQSSGTPAGQGP"
```

```
     misc_difference 2917
                     /note="compared to genome"
                     /replace="a"
ORIGIN

Alignment Scores:
Pred. No.:                4.78e-183       Length:          3498
Score:                    4786.00         Matches:         899
Percent Similarity:       99.89%          Conservative:    1
Best Local Similarity:    99.78%          Mismatches:      1
Query Match:              99.69%          Indels:          0
DB:                       8               Gaps:            0

US-09-920-705-3 (1-901) x AY094463 (1-3498)

Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 314

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 374

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 434

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 494

Qy         81 SerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        495 AGTCAAATTAAGATCGAATACGCACGACCGGCAAAACCTTGTAAGAGTCTATGGGTGGGT 554

Qy        101 GlyIleGlyProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        555 GGAATCGGCCCTAATGTCTCCAAGGATGACCTGGAGGAAGAGTTCAGCAAGTTTGGGAAA 614

Qy        121 IleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMet 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        615 ATCGAGGATTTTAGGTTTCTCAGAGAACGCAAGACAGCTTTCATTGATTATTATGAGATG 674

Qy        141 AspAspAlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArg 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        675 GATGATGCTTTACAGGCTAAGAGCATGAATGGAAAGCCTATGGGTGGTAGCTTTTTGCGT 734

Qy        161 ValAspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsn 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        735 GTTGATTTTCTCCGGTCACAAGCGCCAAAAAAAGAACAATGGGCTGGCTCTTACGATAAC 794

Qy        181 ArgAsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGly 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        795 AGAAATGGCAATATGAATCATAAACCGCAGTATCCTCACTCATATGAAGACTTTAAAGGA 854

Qy        201 AspValGlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 GATGTCCAGCCAAGTAAGGTTCTGTGGATTGGGTTCCCTCCTACTGCTACACAATGCAAT 914

Qy        221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 GATGAGCAAATTCTGCACAATGCGATGATACTCTTTGGTGAGATCGAGAGGGTAAAAAGT 974

Qy        241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 TACCCATCAAGGAATTTTGCACTTGTGGAGTTTAGGAGCGCGGAGGAAGCTCGCCAATGC 1034

Qy        261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 AAGGAAGGCCTACAGGGGAGGTTATTCAATAATCCTAGAATCAAAATTATGTACTCAAAC 1094

Qy        281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       1095 GATGAGTTGCCTCCTGAGCAAGACGATACTAGTTTTTACTCTGGTATGAAACGGTCAAGG 1154

Qy        301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
              ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db       1155 ACAGATATGTTCAATAATGATCCTTCATTTGTATCTTCTCCTCATTCTACTGGAATTCCT 1214

Qy        321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsn 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 GGGTCTATGAGGCCCCTCAGAGGTACAAATGAGCGTTCATATAATGGTGCAGAATACAAT 1274

Qy        341 AspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1275 GACGTTGTTGGTAAGGAGCCAAACTGGAGGAGGCCATCTGCAAATGGAACTGGAATACTC 1334

Qy        361 ProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMet 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1335 CCATCTCCAACAGGACCTGGAATCCTCCCATCTCCTGCACAAGGTACGAGGCGCCCTATG 1394

Qy        381 ArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuValArgGluSerLys 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1395 AGGTCAAACCCCGATTCTTGGGAAGGATATGATCCTGCTCAGTTGGTCAGAGAAAGTAAA 1454

Qy        401 ArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGluArgSer 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1455 CGAACCAGAAGAGATGGATCAGTGGACGGTTTTACTCCAATGGGTGTCGATGAGAGGTCA 1514

Qy        421 PheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAspSerAspHisIle 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1515 TTTGGTCGAGGTTCAGTTGCTGCTAGACCTATCCGTGGCCCCCCTGATTCTGATCACATA 1574

Qy        441 TrpArgGlyMetIleAlaLysGlyGlyThrProValCysCysAlaArgCysValProMet 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1575 TGGAGAGGAATGATTGCCAAGGGTGGAACTCCCGTCTGTTGTGCTCGTTGTGTACCTATG 1634

Qy        461 GlyLysGlyIleGluThrLysLeuProGluValValAsnCysSerAlaArgThrAspLeu 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1635 GGAAAGGGGATTGAAACTAAACTGCCTGAGGTCGTCAATTGTTCAGCAAGAACTGATTTG 1694

Qy        481 AsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGluIleValPhePheValPro 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1695 AATATGCTCGCTAAACATTACGCCGTTGCCATTGGATGTGAGATCGTTTTTTTCGTACCA 1754

Qy        501 AspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAsp 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1755 GACAGGGAAGAAGATTTTGCGTCTTACACTGAATTTCTCCGGTACCTTAGCTCAAAAGAT 1814

Qy        521 ArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProProSerAsp 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1815 CGGGCGGGTGTTGCCAAATTAGATGATGGTACAACTTTATTCTTGGTGCCTCCATCAGAT 1874

Qy        541 PheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLys 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1875 TTCTTAACTGATGTACTCCAAGTGACCCGTCAAGAACGGCTATATGGTGTTGTTCTCAAG 1934

Qy        561 LeuProProProAlaValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnPro 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1935 TTACCCCCGCCAGCCGTTCCTGTTACAGCATCATACAGACAAGAATCTCAGTCCAATCCT 1994

Qy        581 LeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrPro 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1995 CTGCATTATATGGATCAAGCCCGGGATTCACCTGCCAATGCTAGTCACAGTTTATATCCT 2054

Qy        601 ProArgGluAsnTyrIleArgGlyAlaProGluHisLeuThrAlaAlaSerLysProSer 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2055 CCTAGGGAAAATTACATTAGGGGTGCACCAGAACATTTGACAGCTGCTTCAAAACCATCT 2114

Qy        621 ValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThr 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2115 GTTAGCGAGCCTCTCAGAATACCTAATAATGCAGCGCCTCAAGCTGGGGTTAGTTTAACT 2174

Qy        641 ProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2175 CCGGAGCTTTTAGCCACTCTGGCATCTATTCTCCCTGCAACTTCTCAACCTGCTGCCCCT 2234
```

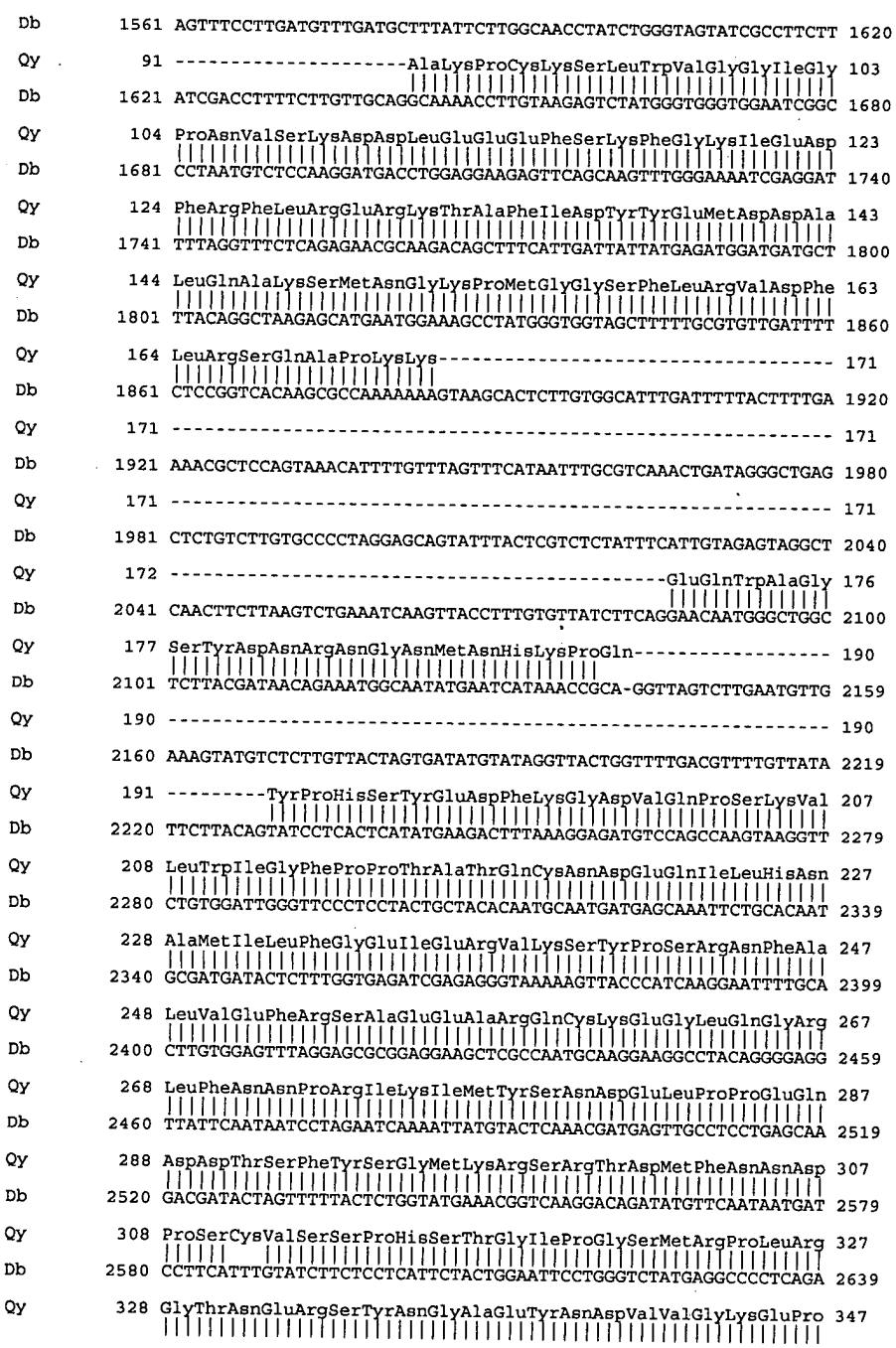
```
Qy        661 GluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsn 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2235 GAGAGTCACCAACCTATGTCAGGACCTTCAACAGTTGTTTCCACAGCACATCAGTCCAAT 2294

Qy        681 GlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHis 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2295 GGACTGTACAATGGAGAAGCACCGTCTCAAGCTTGGAAAAGAGGTCCACAAACAGTTCAT 2354

Qy        701 AspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeu 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2355 GATGCGTCAAATCAGTCATTCCAACAATACGGAAATCAGTACACTCCAGCTGGGCAACTA 2414

Qy        721 ProProProProSerArgTyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMet 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2415 CCTCCTCCTCCTTCGCGTTACCCTCCAGCTTCAAACAACCCCAACTACACTAGTGGAATG 2474

Qy        741 ValHisGlyAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeu 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2475 GTCCATGGCAACATGCAATACCAGAGCCAATCTGTTAACATGCCTCAGCTGTCTCCGTTA 2534

Qy        761 ProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProVal 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2535 CCAAATATGCCTCATAATAATTATTCCATGTACACTCAGGGTTCGTCAAATCATCCTGTT 2594

Qy        781 SerGlnProMetValGlnGlnTyrGlnProGluAlaSerMetProAsnGlnAsnTyrGly 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2595 TCTCAGCCCATGGTCCAGCAATACCAACCAGAAGCGTCCATGCCAAACCAAAACTATGGT 2654

Qy        801 ProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsn 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2655 CCAATTCCAAGTTATCAGCAAGCTAATTTTCATGGCGTAACAACAAATCAGGCACAGAAC 2714

Qy        821 LeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGlu 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2715 TTAAACCCTTCCCAATTTCAAGCTGCCATGCAACCACCAGCAGATAAGGCAAATTTAGAG 2774

Qy        841 ProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThr 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2775 CCACAAAACCAAGCACTACGATTGCAGCCTATGATCTCTGGGGATGGTCAGGGTACAACA 2834

Qy        861 AspGlyGluValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsnLeu 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2835 GATGGGGAGGTCGATAAGAATCAGAGATACCAGTCAACACTACAATTTGCAGCAAACCTT 2894

Qy        881 LeuLeuGlnIleGlnGlnLysGlnGlnGlnGlnSerSerGlyThrProAlaGlyGlnGly 900
              ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db       2895 CTTCTCCAGATACAGCAGAAACGGCAGCAACAGTCTTCAGGTACTCCGGCTGGACAGGGG 2954

Qy        901 Pro 901
              |||
Db       2955 CCT 2957
```

```
RESULT 4
CQ804588
LOCUS       CQ804588               3171 bp    DNA     linear   PAT 10-MAY-2004
DEFINITION  Sequence 999 from Patent WO2004035798.
ACCESSION   CQ804588
VERSION     CQ804588.1  GI:47110928
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Inze,D., de Veylder,L. and Vlieghe,K.
  TITLE     Identification of novel e2f target genes and use thereof
  JOURNAL   Patent: WO 2004035798-A 999 29-APR-2004;
            CropDesign N.V. (BE)
FEATURES             Location/Qualifiers
     source          1. .3171
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
ORIGIN      .
```

```
Alignment Scores:
Pred. No.:              6.6e-178        Length:         3171
Score:                  4656.50         Matches:        892
Percent Similarity:     85.04%          Conservative:   6
Best Local Similarity:  84.47%          Mismatches:     3
Query Match:            96.99%          Indels:         155
DB:                     6               Gaps:           3

US-09-920-705-3 (1-901) x CQ804588 (1-3171)

Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 60

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 120

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 180

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 240

Qy         81 Ser---------------------------------------------------------- 81
              :::
Db        241 ACGAATGAGCTCTGCAAAAGAATACATCAAACACATGTTTACTCTCGTTTATGCGAAGAA 300

Qy         81 ---------------------------------------------------------- 81

Db        301 GGTTTAATCCGATTGTCGTTTATCTGCAAACTTGTATCCGCGTTTAGCAGTCTGATTTCA 360

Qy         81 ---------------------------------------------------------- 81

Db        361 GGTTCGCTTCAGATGTTAAATCTCACAAGCTTGAGTATGAATGTATTGCGTACTCCACTC 420

Qy         82 --------------------------GlnIleLysIleGluTyr--------------- 87
                                         :::|||:::::::::
Db        421 GAAATTCGCAATGATAAGAAATTTTGGAGTTTAAAAGTTCAGTTTCCTGCAAAAATCAAA 480

Qy         87 ---------------------------------------------------------- 87

Db        481 CCGCGTGGAGAAATTTGTCTATGGCGGTACGTAGATATCAATAACTGCATCTGCGAAACA 540

Qy         87 ---------------------------------------------------------- 87

Db        541 GAAAATTCTGGCAACCGTATCGTCCTAACTCCTAGTATCGCTGGCACATATCCATATCAG 600

Qy         87 ---------------------------------------------------------- 87

Db        601 TGTGAGGGAGTTTTGGGCTTGTTTTGGCTCGATGCTTCAGGAAGACAAATTACGTGGTTG 660

Qy         87 ---------------------------------------------------------- 87

Db        661 TTAAGGCGGCTAACTCTACCAATCAGAAACGCTTTATTCGAAGAACCATGTTTGTTCCTC 720

Qy         88 ---AlaArgPro---AlaLysProCysLysSerLeuTrpValGlyGlyIleGlyProAsn 105
                  :::   |||   ||||||||||||||||||||||||||||||||||||||||||||
Db        721 AATTCCCATCCCTACGCAAAACCTTGTAAGAGTCTATGGGTGGGTGGAATCGGCCCTAAT 780

Qy        106 ValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArg 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTCTCCAAGGATGACCTGGAGGAAGAGTTCAGCAAGTTTGGGAAAATCGAGGATTTTAGG 840

Qy        126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTCTCAGAGAACGCAAGACAGCTTTCATTGATTATTATGAGATGGATGATGCTTTACAG 900
```

```
Qy      146 AlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeuArg 165
Db      901 GCTAAGAGCATGAATGGAAAGCCTATGGGTGGTAGCTTTTTGCGTGTTGATTTTCTCCGG 960

Qy      166 SerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsnGlyAsnMet 185
Db      961 TCACAAGCGCCAAAAAAAGAACAATGGGCTGGCTCTTACGATAACAGAAATGGCAATATG 1020

Qy      186 AsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspValGlnProSer 205
Db      1021 AATCATAAACCGCAGTATCCTCACTCATATGAAGACTTTAAAGGAGATGTCCAGCCAAGT 1080

Qy      206 LysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeu 225
Db      1081 AAGGTTCTGTGGATTGGGTTCCCTCCTACTGCTACACAATGCAATGATGAGCAAATTCTG 1140

Qy      226 HisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsn 245
Db      1141 CACAATGCGATGATACTCTTTGGTGAGATCGAGAGGGTAAAAAGTTACCCATCAAGGAAT 1200

Qy      246 PheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeuGln 265
Db      1201 TTTGCACTTGTGGAGTTTAGGAGCGCGGAGGAAGCTCGCCAATGCAAGGAAGGCCTACAG 1260

Qy      266 GlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProPro 285
Db      1261 GGGAGGTTATTCAATAATCCTAGAATCAAAATTATGTACTCAAACGATGAGTTGCCTCCT 1320

Qy      286 GluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsn 305
Db      1321 GAGCAAGACGATACTAGTTTTTACTCTGGTATGAAACGGTCAAGGACAGATATGTTCAAT 1380

Qy      306 AsnAspProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArgPro 325
Db      1381 AATGATCCTTCATTTGTATCTTCTCCTCATTCTACTGGAATTCCTGGGTCTATGAGGCCC 1440

Qy      326 LeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValValGlyLys 345
Db      1441 CTCAGAGGTACAAATGAGCGTTCATATAATGGTGCAGAATACAATGACGTTGTTGGTAAG 1500

Qy      346 GluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGly 365
Db      1501 GAGCCAAACTGGAGGAGGCCATCTGCAAATGGAACTGGAATACTCCCATCTCCAACAGGA 1560

Qy      366 ProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAsp 385
Db      1561 CCTGGAATCCTCCCATCTCCTGCACAAGGTACGAGGCGCCCTATGAGGTCAAACCCCGAT 1620

Qy      386 SerTrpGluGlyTyrAspProAlaGlnLeuValArgGluSerLysArgThrArgArgAsp 405
Db      1621 TCTTGGGAAGGATATGATCCTGCTCAGTTGGTCAGAGAAAGTAAACGAACCAGAAGAGAT 1680

Qy      406 GlySerValAspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGlySer 425
Db      1681 GGATCAGTGGACGGTTTTACTCCAATGGGTGTCGATGAGAGGTCATTTGGTCGAGGTTCA 1740

Qy      426 ValAlaAlaArgProIleArgGlyProProAspSerAspHisIleTrpArgGlyMetIle 445
Db      1741 GTTGCTGCTAGACCTATCCGTGGCCCCCCTGATTCTGATCACATATGGAGAGGAATGATT 1800

Qy      446 AlaLysGlyGlyThrProValCysCysAlaArgCysValProMetGlyLysGlyIleGlu 465
Db      1801 GCCAAGGGTGGAACTCCCGTCTGTTGTGCTCGTTGTGTACCTATGGGAAAGGGGATTGAA 1860

Qy      466 ThrLysLeuProGluValValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLys 485
Db      1861 ACTAAACTGCCTGAGGTCGTCAATTGTTCAGCAAGAACTGATTTGAATATGCTCGCTAAA 1920

Qy      486 HisTyrAlaValAlaIleGlyCysGluIleValPhePheValProAspArgGluGluAsp 505
Db      1921 CATTACGCCGTTGCCATTGGATGTGAGATCGTTTTTTTCGTACCAGACAGGGAAGAAGAT 1980

Qy      506 PheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAla 525
Db      1981 TTTGCGTCTTACACTGAATTTCTCCGGTACCTTAGCTCAAAAGATCGGGCGGGTGTTGCC 2040

Qy      526 LysLeuAspAspGlyThrThrLeuPheLeuValProProSerAspPheLeuThrAspVal 545
Db      2041 AAATTAGATGATGGTACAACTTTATTCTTGGTGCCTCCATCAGATTTCTTAACTGATGTA 2100

Qy      546 LeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLysLeuProProProAla 565
Db      2101 CTCCAAGTGACCCGTCAAGAACGGCTATATGGTGTTGTTCTCAAGTTACCCCCGCCAGCC 2160

Qy      566 ValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnProLeuHisTyrMetAsp 585
Db      2161 GTTCCTGTTACAGCATCATACAGACAAGAATCTCAGTCCAATCCTCTGCATTATATGGAT 2220

Qy      586 GlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGluAsnTyr 605
Db      2221 CAAGCCCGGGATTCACCTGCCAATGCTAGTCACAGTTTATATCCTCCTAGGGAAAATTAC 2280

Qy      606 IleArgGlyAlaProGluHisLeuThrAlaAlaSerLysProSerValSerGluProLeu 625
Db      2281 ATTAGGGGTGCACCAGAACATTTGACAGCTGCTTCAAAACCATCTGTTAGCGAGCCTCTC 2340

Qy      626 ArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAla 645
Db      2341 AGAATACCTAATAATGCAGCGCCTCAAGCTGGGGTTAGTTTAACTCCGGAGCTTTTAGCC 2400

Qy      646 ThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHisGlnPro 665
Db      2401 ACTCTGGCATCTATTCTCCCTGCAACTTCTCAACCTGCTGCCCCTGAGAGTCACCAACCT 2460

Qy      666 MetSerGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGly 685
Db      2461 ATGTCAGGACCTTCAACAGTTGTTTCCACAGCACATCAGTCCAATGGACTGTACAATGGA 2520

Qy      686 GluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHisAspAlaSerAsnGln 705
Db      2521 GAAGCACCGTCTCAAGCTTGGAAAAGAGGTCCACAAACAGTTCATGATGCGTCAAATCAG 2580

Qy      706 SerPheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeuProProProProSer 725
Db      2581 TCATTCCAACAATACGGAAATCAGTACACTCCAGCTGGGCAACTACCTCCTCCTCCTTCG 2640

Qy      726 ArgTyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMet 745
Db      2641 CGTTACCCTCCAGCTTCAAACAACCCCAACTACACTAGTGGAATGGTCCATGGCAACATG 2700

Qy      746 GlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeuProAsnMetProHis 765
Db      2701 CAATACCAGAGCCAATCTGTTAACATGCCTCAGCTGTCTCCGTTACCAAATATGCCTCAT 2760

Qy      766 AsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProValSerGlnProMetVal 785
Db      2761 AATAATTATTCCATGTACACTCAGGGTTCGTCAAATCATCCTGTTTCTCAGCCCATGGTC 2820

Qy      786 GlnGlnTyrGlnProGluAlaSerMetProAsnGlnAsnTyrGlyProIleProSerTyr 805
Db      2821 CAGCAATACCAACCAGAAGCGTCCATGCCAAACCAAAACTATGGTCCAATTCCAAGTTAT 2880

Qy      806 GlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGln 825
Db      2881 CAGCAAGCTAATTTTCATGGCGTAACAACAAATCAGGCACAGAACTTAAACCCTTCCCAA 2940

Qy      826 PheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGluProGlnAsnGlnAla 845
Db      2941 TTTCAAGCTGCCATGCAACCACCAGCAGATAAGGCAAATTTAGAGCCACAAAACCAAGCA 3000

Qy      846 LeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThrAspGlyGluValAsp 865
Db      3001 CTACGATTGCAGCCTATGATCTCTGGGGATGGTCAGGGTACAACAGATGGGGAGGTCGAT 3060

Qy      866 LysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsnLeuLeuLeuGlnIleGln 885
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 AAGAATCAGAGATACCAGTCAACACTACAATTTGCAGCAAACCTTCTTCTCCAGATACAG 3120

Qy        886 GlnLysGlnGlnGlnGlnSerSerGlyThrProAlaGlyGlnGlyPro 901
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 CAGAAACAGCAGCAACAGTCTTCAGGTACTCCGGCTGGACAGGGGCCT 3168

RESULT 5
AX420419
LOCUS       AX420419                4593 bp    DNA     linear   PAT 18-JUN-2002
DEFINITION  Sequence 1 from Patent WO0212518.
ACCESSION   AX420419
VERSION     AX420419.1  GI:21524575
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Amasino,R.M., Schomburg,F.M., Michaels,S.D. and Patton,D.
  TITLE     Floral induction gene
  JOURNAL   Patent: WO 0212518-A 1 14-FEB-2002;
            WISCONSIN ALUMNI RESEARCH FOUNDATION (US)
FEATURES             Location/Qualifiers
     source          1. .4593
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
ORIGIN

Alignment Scores:
Pred. No.:              3.81e-167       Length:         4593
Score:                  4392.00         Matches:        899
Percent Similarity:     58.72%          Conservative:   0
Best Local Similarity:  58.72%          Mismatches:     2
Query Match:            91.48%          Indels:         631
DB:                     6               Gaps:           4

US-09-920-705-3 (1-901) x AX420419 (1-4593)

Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 60

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 120

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 180

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 240

Qy         81 SerGlnIleLysIleGluTyrAlaArgPro---------------------------- 90
              ||||||||||||||||||||||||||||
Db        241 AGTCAAATTAAGATCGAATACGCACGACCGGTTTGTTCTTATCTATATCTTCGTTTGTTC 300

Qy         90 ------------------------------------------------------------ 90

Db        301 TCTAACTTTGATTGTCTTTTGTCAACGATTATACTCTTTTTGCGAATTCATAGTCCAGGT 360

Qy         90 ------------------------------------------------------------ 90

Db        361 TCACAAAACTTTGATGATGCTTGTTTAGTCCAAAAATTTCTTGTTGAATCTGTTTTTTTT 420

Qy         90 ------------------------------------------------------------ 90

Db        421 TTCCTCATCGTACAAATCAAAGTCGAAACCTAGTTTTTTTCTATTATACGTCGTTAGCTT 480

Qy         90 ------------------------------------------------------------ 90

Db        481 AAGGCGAAACCTGATCCGATCGAAACGTCTTTTCTCAAATTACTTTGGTTATATCGAACT 540

Qy         90 ------------------------------------------------------------ 90

Db        541 CGCGCAAAGCCAAACCACAGAGAAGCTCTGCAAAATTTGATGTTAAAGCATATATAACTC 600

Qy         90 ------------------------------------------------------------ 90

Db        601 TTAGCGAATGAGCTCTGCAAAAGAATACATCAAACACATGTTTACTCTCGTTTATGCGAA 660

Qy         90 ------------------------------------------------------------ 90

Db        661 GAAGGTTTAATCCGATTGTCGTTTATCTGCAAACTTGTATCCGCGTTTAGCAGTCTGATT 720

Qy         90 ------------------------------------------------------------ 90

Db        721 TCAGGTTCGCTTCAGATGTTAAATCTCACAAGCTTGAGTATGAATGTATTGCGTACTCCA 780

Qy         90 ------------------------------------------------------------ 90

Db        781 CTCGAAATTCGCAATGGTGAAAATTTGGAAGCGAGCAAACTCTATCATCGGCCAAACAGA 840

Qy         90 ------------------------------------------------------------ 90

Db        841 TAAGAAATTTTGGAGTTTAAAAGTTCAGTTTCCTGCAAAAATCAAACCGCGTGGAGAAAT 900

Qy         90 ------------------------------------------------------------ 90

Db        901 TTGTCTATGGCGGTACGTAGATATCAATAACTGCATCTGCGAAACAGAAAATTCTGGCAA 960

Qy         90 ------------------------------------------------------------ 90

Db        961 CCGTATCGTCCTAACTCCTAGTATCGCTGGCACATATCCATATCAGTGTGAGGGAGTTTT 1020

Qy         90 ------------------------------------------------------------ 90

Db       1021 GGGCTTGTTTTGGCTCGATGCTTCAGGAAGACAAATTACGTGGTTGTTAAGGCGGCTAAC 1080

Qy         90 ------------------------------------------------------------ 90

Db       1081 TCTACCAATCAGAAACGCTTTATTCGAAGAACCATGTTTGTTCCTCAATTCCCATCCCTA 1140

Qy         90 ------------------------------------------------------------ 90

Db       1141 CGTACAATCTGGGCTTTCCCATTGTAGTTCCTTAGGAAGTTGACTTCTTCACAACCATTC 1200

Qy         90 ------------------------------------------------------------ 90

Db       1201 TGGGATTGGTACAGTTGCAGTGGACAAGACAAATATTCATTTGCAGGCAGACTCAAATTT 1260

Qy         90 ------------------------------------------------------------ 90

Db       1261 CAATGTCTTGCCGGCTTTGAGTACTTAAATGGAGTGTTCAGGGATTGGTTTATTTGGGAC 1320

Qy         90 ------------------------------------------------------------ 90

Db       1321 TCAGGCAATGGATAAAGACAGGAATGTTTTGCAGAAAGTATTAATGTCTTTCCGGCTTTG 1380

Qy         90 ------------------------------------------------------------ 90

Db       1381 GGGACTACATACTTCTCCTACAGACAAAATCTGATGTTTAGGCAAGAGAAACTATTAACA 1440

Qy         90 ------------------------------------------------------------ 90

Db       1441 CTGATTTAATAGAGAAAGAGGAGATGGTTTCTCCTTGCGGCAATTTTATTTGTTTAGGAA 1500

Qy         90 ------------------------------------------------------------ 90

Db       1501 AGCAATTGATATGAATTGGTGTCGTAGTGTAGTTGAAATTACTAGTTAGTTTGTGTGTTT 1560

Qy         90 ------------------------------------------------------------ 90
```


Db 1561 AGTTTCCTTGATGTTTGATGCTTTATTCTTGGCAACCTATCTGGGTAGTATCGCCTTCTT 1620
Qy 91 ----------AlaLysProCysLysSerLeuTrpValGlyGlyIleGly 103
Db 1621 ATCGACCTTTTCTTGTTGCAGGCAAAACCTTGTAAGAGTCTATGGGTGGGTGGAATCGGC 1680
Qy 104 ProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAsp 123
Db 1681 CCTAATGTCTCCAAGGATGACCTGGAGGAAGAGTTCAGCAAGTTTGGGAAAATCGAGGAT 1740
Qy 124 PheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAla 143
Db 1741 TTTAGGTTTCTCAGAGAACGCAAGACAGCTTTCATTGATTATTATGAGATGGATGATGCT 1800
Qy 144 LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPhe 163
Db 1801 TTACAGGCTAAGAGCATGAATGGAAAGCCTATGGGTGGTAGCTTTTTGCGTGTTGATTTT 1860
Qy 164 LeuArgSerGlnAlaProLysLys---------- 171
Db 1861 CTCCGGTCACAAGCGCCAAAAAAAGTAAGCACTCTTGTGGCATTTGATTTTTACTTTTGA 1920
Qy 171 ---------- 171
Db 1921 AAACGCTCCAGTAAACATTTTGTTTAGTTTCATAATTTGCGTCAAACTGATAGGGCTGAG 1980
Qy 171 ---------- 171
Db 1981 CTCTGTCTTGTGCCCCTAGGAGCAGTATTTACTCGTCTCTATTTCATTGTAGAGTAGGCT 2040
Qy 172 ----------GluGlnTrpAlaGly 176
Db 2041 CAACTTCTTAAGTCTGAAATCAAGTTACCTTTGTGTTATCTTCAGGAACAATGGGCTGGC 2100
Qy 177 SerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGln---------- 190
Db 2101 TCTTACGATAACAGAAATGGCAATATGAATCATAAACCGCA-GGTTAGTCTTGAATGTTG 2159
Qy 190 ---------- 190
Db 2160 AAAGTATGTCTCTTGTTACTAGTGATATGTATAGGTTACTGGTTTTGACGTTTTGTTATA 2219
Qy 191 ---------TyrProHisSerTyrGluAspPheLysGlyAspValGlnProSerLysVal 207
Db 2220 TTCTTACAGTATCCTCACTCATATGAAGACTTTAAAGGAGATGTCCAGCCAAGTAAGGTT 2279
Qy 208 LeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHisAsn 227
Db 2280 CTGTGGATTGGGTTCCCTCCTACTGCTACACAATGCAATGATGAGCAAATTCTGCACAAT 2339
Qy 228 AlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsnPheAla 247
Db 2340 GCGATGATACTCTTTGGTGAGATCGAGAGGGTAAAAAGTTACCCATCAAGGAATTTTGCA 2399
Qy 248 LeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeuGlnGlyArg 267
Db 2400 CTTGTGGAGTTTAGGAGCGCGGAGGAAGCTCGCCAATGCAAGGAAGGCCTACAGGGGAGG 2459
Qy 268 LeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProProGluGln 287
Db 2460 TTATTCAATAATCCTAGAATCAAAATTATGTACTCAAACGATGAGTTGCCTCCTGAGCAA 2519
Qy 288 AspAspThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsnAsp 307
Db 2520 GACGATACTAGTTTTTACTCTGGTATGAAACGGTCAAGGACAGATATGTTCAATAATGAT 2579
Qy 308 ProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArgProLeuArg 327
Db 2580 CCTTCATTTGTATCTTCTCCTCATTCTACTGGAATTCCTGGGTCTATGAGGCCCCTCAGA 2639
Qy 328 GlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValValGlyLysGluPro 347
Db 2640 GGTACAAATGAGCGTTCATATAATGGTGCAGAATACAATGACGTTGTTGGTAAGGAGCCA 2699
Qy 348 AsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGly 367
Db 2700 AACTGGAGGAGGCCATCTGCAAATGGAACTGGAATACTCCCATCTCCAACAGGACCTGGA 2759
Qy 368 IleLeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrp 387
Db 2760 ATCCTCCCATCTCCTGCACAAGGTACGAGGCGCCCTATGAGGTCAAACCCCGATTCTTGG 2819
Qy 388 GluGlyTyrAspProAlaGlnLeuValArgGluSerLysArgThrArgArgAspGlySer 407
Db 2820 GAAGGATATGATCCTGCTCAGTTGGTCAGAGAAAGTAAACGAACCAGAAGAGATGGATCA 2879
Qy 408 ValAspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGlySerValAla 427
Db 2880 GTGGACGGTTTTACTCCAATGGGTGTCGATGAGAGGTCATTTGGTCGAGGTTCAGTTGCT 2939
Qy 428 AlaArgProIleArgGlyProProAspSerAspHisIleTrpArgGlyMetIleAlaLys 447
Db 2940 GCTAGACCTATCCGTGGCCCCCCTGATTCTGATCACATATGGAGAGGAATGATTGCCAAG 2999
Qy 448 GlyGlyThrProValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLys 467
Db 3000 GGTGGAACTCCCGTCTGTTGTGCTCGTTGTGTACCTATGGGAAAGGGGATTGAAACTAAA 3059
Qy 468 Leu---------- 468
Db 3060 CTGTGAGTACTAATTTCTAGCACTTTAACCCTTCTAGTGTTTTCTTTTTCAGAGCGATTT 3119
Qy 468 ---------- 468
Db 3120 ATATATTTTCCATTTCATTCTCGATGGAAGTAACATTATTATAGATAGTACATTTTTATT 3179
Qy 468 ---------- 468
Db 3180 TTACTATTACTTGTTTAGTTTCTGAGATGTCTTGATTTTCATGGTGTTGATTCATTTTTG 3239
Qy 469 ----------ProGluVa 471
Db 3240 GCATTGCCCTCAATTACTGACTTTGTTTTTTTTTTAATAATTGATTTATAGGCCTGAGGT 3299
Qy 471 lValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIl 491
Db 3300 CGTCAATTGTTCAGCAAGAACTGATTTGAATATGCTCGCTAAACATTACGCCGTTGCCAT 3359
Qy 491 eGlyCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGl 511
Db 3360 TGGATGTGAGATCGTTTTTTTCGTACCAGACAGGGAAGAAGATTTTGCGTCTTACACTGA 3419
Qy 511 uPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyTh 531
Db 3420 ATTTCTCCGGTACCTTAGCTCAAAAGATCGGGCGGGTGTTGCCAAATTAGATGATGGTAC 3479
Qy 531 rThrLeuPheLeuValProProSerAspPheLeuThrAspValLeuGlnValThrArgGl 551
Db 3480 AACTTTATTCTTGGTGCCTCCATCAGATTTCTTAACTGATGTACTCCAAGTGACCCGTCA 3539
Qy 551 nGluArgLeuTyrGlyValValLeuLysLeuProProProAlaValProValThrAlaSe 571
Db 3540 AGAACGGCTATATGGTGTTGTTCTCAAGTTACCCCCGCCAGCCGTTCCTGTTACAGCATC 3599
Qy 571 rTyrArgGlnGluSerGlnSerAsnProLeuHisTyrMetAspGlnAlaArgAspSerPr 591
Db 3600 ATACAGACAAGAATCTCAGTCCAATCCTCTGCATTATATGGATCAAGCCCGGGATTCACC 3659
Qy 591 oAlaAsnAlaSerHisSerLeuTyrProProArgGluAsnTyrIleArgGlyAlaProGl 611
Db 3660 TGCCAATGCTAGTCACAGTTTATATCCTCCTAGGGAAAATTACATTAGGGGTGCACCAGA 3719
Qy 611 uHisLeuThrAlaAlaSerLysProSerValSerGluProLeuArgIleProAsnAsnAl 631
Db 3720 ACATTTGACAGCTGCTTCAAAACCATCTGTTAGCGAGCCTCTCAGAATACCTAATAATGC 3779

```
Qy        631 aAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAlaThrLeuAlaSerIleLe 651
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3780 AGCGCCTCAAGCTGGGGTTAGTTTAACTCCGGAGCTTTTAGCCACTCTGGCATCTATTCT 3839

Qy        651 uProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMetSerGlyProSerTh 671
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3840 CCCTGCAACTTCTCAACCTGCTGCCCCTGAGAGTCACCAACCTATGTCAGGACCTTCAAC 3899

Qy        671 rValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyGluAlaProSerGlnAl 691
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3900 AGTTGTTTCCACAGCACATCAGTCCAATGGACTGTACAATGGAGAAGCACCGTCTCAAGC 3959

Qy        691 aTrpLysArgGlyProGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTyrGl 711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3960 TTGGAAAAGAGGTCCACAAACAGTTCATGATGCGTCAAATCAGTCATTCCAACAATACGG 4019

Qy        711 yAsnGlnTyrThrProAlaGlyGlnLeuProProProProSerArgTyrProProAlaSe 731
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4020 AAATCAGTACACTCCAGCTGGGCAACTACCTCCTCCTCCTTCGCGTTACCCTCCAGCTTC 4079

Qy        731 rAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMetGlnTyrGlnSerGlnSe 751
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4080 AAACAACCCCAACTACACTAGTGGAATGGTCCATGGCAACATGCAATACCAGAGCCAATC 4139

Qy        751 rValAsnMetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyrSerMetTy 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4140 TGTTAACATGCCTCAGCTGTCTCCGTTACCAAATATGCCTCATAATAATTATTCCATGTA 4199

Qy        771 rThrGlnGlySerSerAsnHisProValSerGlnProMetValGlnGlnTyrGlnProGl 791
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4200 CACTCAGGGTTCGTCAAATCATCCTGTTTCTCAGCCCATGGTCCAGCAATACCAACCAGA 4259

Qy        791 uAlaSerMetProAsnGlnAsnTyrGlyProIleProSerTyrGlnGlnAlaAsnPheHi 811
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4260 AGCGTCCATGCCAAACCAAAACTATGGTCCAATTCCAAGTTATCAGCAAGCTAATTTTCA 4319

Qy        811 sGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGl 831
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4320 TGGCGTAACAACAAATCAGGCACAGAACTTAAACCCTTCCCAATTTCAAGCTGCCATGCA 4379

Qy        831 nProProAlaAspLysAlaAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMe 851
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4380 ACCACCAGCAGATAAGGCAAATTTAGAGCCACAAAACCAAGCACTACGATTGCAGCCTAT 4439

Qy        851 tIleSerGlyAspGlyGlnGlyThrThrAspGlyGluValAspLysAsnGlnArgTyrGl 871
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4440 GATCTCTGGGGATGGTCAGGGTACAACAGATGGGGAGGTCGATAAGAATCAGAGATACCA 4499

Qy        871 nSerThrLeuGlnPheAlaAlaAsnLeuLeuLeuGlnIleGlnGlnLysGlnGlnGlnGl 891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4500 GTCAACACTACAATTTGCAGCAAACCTTCTTCTCCAGATACAGCAGAAACAGCAGCAACA 4559

Qy        891 nSerSerGlyThrProAlaGlyGlnGlyPro 901
              |||||||||||||||||||||||||||||
Db       4560 GTCTTCAGGTACTCCGGCTGGACAGGGGCCT 4590
```

```
RESULT 6
AC002335/c
LOCUS       AC002335               102057 bp    DNA     linear   PLN 11-MAR-2002
DEFINITION  Arabidopsis thaliana chromosome 2 clone T1O24 map CIC10F02,
            complete sequence.
ACCESSION   AC002335
VERSION     AC002335.3  GI:20196857
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 102057)
  AUTHORS   Rounsley,S.D., Tschudy,M.M., Lin,X., Ketchum,K.A., Crosby,M.L.,
            Brandon,R.C., Spriggs,T.A., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
            Somerville,C.R. and Venter,J.C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 102057)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 102057)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Apr 18, 2002 this sequence version replaced gi:6598343.
FEATURES             Location/Qualifiers
     source          1. .102057
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /map="CIC10F02"
                     /clone="T1O24"
     misc_feature    1. .16126
                     /note="overlap with BAC clone F14B2
                     (AC004450:74158. .90283)."
     repeat_region   complement(334. .355)
                     /rpt_family="AT_rich"
     gene            complement(3113. .4397)
                     /gene="At2g43270"
                     /note="synonym: T1O24.1"
     mRNA            complement(join(<3113. .3691,3734. .3924,4004. .>4397))
                     /gene="At2g43270"
     CDS             complement(join(3113. .3691,3734. .3924,4004. .4397))
                     /gene="At2g43270"
                     /codon_start=1
                     /product="putative myb-related transcription factor"
                     /protein_id="AAB64309.2"
                     /db_xref="GI:20196861"
                     /translation="MIYVVPDLLEEIFLGLPLKSILRFKTVSKQWRSILESKSFAERR
                     LNVEKKEKILAVGDRTELGFEGEEEIQMVYLHCDIDATRPSLTCEGLVCIPAPGWINV
                     LNPSTRQLRRFPCSPNHHVPSDRIRFQFRDELYLTSFPGNWEMGFGRDKFNGSYKVVR
                     MCFSPVEKCEVLDVETGEWSELNPPPNDIDVGRKSRFGILALDLHTLEFHNVSVPSTC
                     CTMSAQMVNLDDRLAITKTNGAPEWELEILSLMDVDKEIWSKTYSITLASIIINHPWE
                     SSWFTVLTVSKLGNLVFCDNHKRLFKYNTETNEICCLSSDISVISPYLEDLVPLRLDS
                     GHDHRDYKIRTSSCGLFSKHPEAGLFKEIITFRFVFATLVLVGYQYYRSLKFK"
     gene            5148. .5767
                     /gene="At2g43280"
                     /note="synonym: T1O24.2; predicted by genscan"
     mRNA            join(<5148. .5447,5552. .>5767)
                     /gene="At2g43280"
     CDS             join(5148. .5447,5552. .5767)
                     /gene="At2g43280"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAB64334.1"
                     /db_xref="GI:2289005"
                     /translation="MFYDDYSRRLGFVMRVMSCRRSEKDGRILARRFGCNKEGHCVSI
                     RGKFGSVRKPRPSTREGCKAMIHVKYDRSGKWVITKFVKEHNHPLVVSPREARHTLDE
                     KDKRIQELTIELRNKKRLCAAYKEQLDAFAKIVEEHSNQIAKKVENVVNNLKEFEHLE
                     HALLRSKQDAI"
     gene            complement(5982. .6953)
                     /gene="At2g43290"
                     /note="synonym: T1O24.3; supported by full length cDNA:
                     Ceres:31535"
     mRNA            complement(<5982. .>6953)
                     /gene="At2g43290"
     CDS             complement(6214. .6861)
                     /gene="At2g43290"
                     /codon_start=1
                     /product="putative calcium binding protein"
                     /protein_id="AAB64310.2"
                     /db_xref="GI:20196862"
                     /translation="MVRIPLLYNILNSFLLSLVPKKLRTLFPLSWFDKTLHKNSPPSP
```

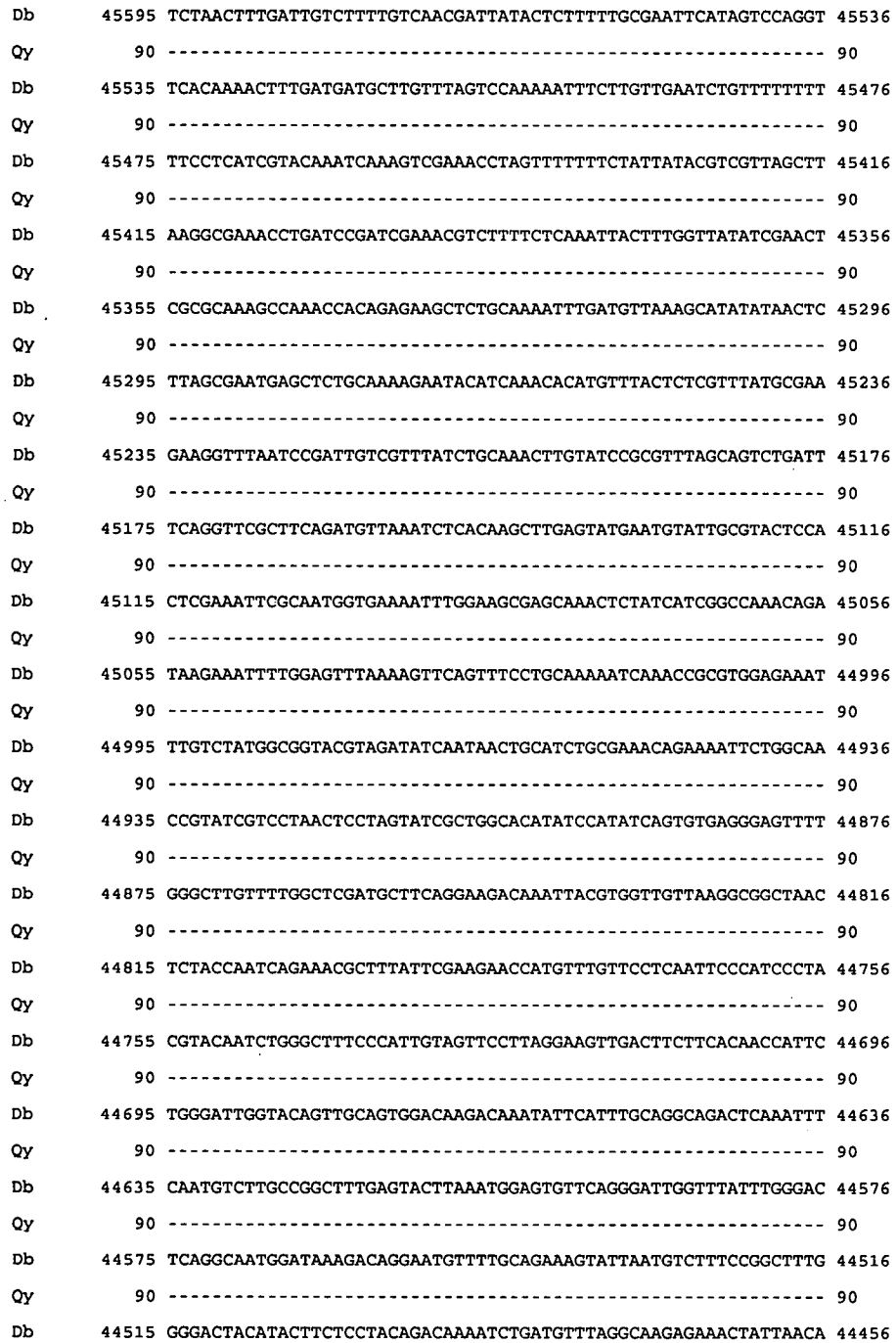
```
                STMLPSPSSSSAPTKRIDPSELKRVFQMFDKNGDGRITKEELNDSLENLGIYIPDKDL
                TQMIHKIDANGDGCVDIDEFESLYSSIVDEHHNDGETEEEDMKDAFNVFDQDGDGFIT
                VEELKSVMASLGLKQGKTLDGCKKMIMQVDADGDGRVNYKEFLQMMKGGGFSSSN"
repeat_region   complement(6998. .7029)
                /rpt_family="AT_rich"
repeat_region   complement(7486. .7547)
                /rpt_family="AT_rich"
repeat_region   complement(7628. .7660)
                /rpt_family="AT_rich"
gene            complement(10479. .10550)
                /gene="At2g43300"
tRNA            complement(10479. .10550)
                /gene="At2g43300"
                /product="tRNA-Thr"
gene            11012. .11413
                /gene="At2g43310"
                /note="synonym: T1O24.5; predicted by genscan"
mRNA            <11012. .>11413
                /gene="At2g43310"
CDS             11012. .11413
                /gene="At2g43310"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAB64333.1"
                /db_xref="GI:2289004"
                /translation="MNGTHRHHLLQLVLSCRKITAQVTQPGSSTIIAMASSSEQEFLA
                KSRANLYRFPHSNHFWDSKTASRVGEKLGLRLRELGVDVVSIDADEEISRPIHHRKRV
                LPLFDSVRRTGIRVDGTEQLNDISVPIIPKN"
gene            complement(11649. .14833)
                /gene="At2g43320"
                /note="synonym: T1O24.6; predicted by genscan; supported
                by cDNA: gi_14517475_gb_AY039573 1_"
mRNA            complement(join(<11649. .11950,12040. .12519,12610. .12654,
                13002. .13102,13191. .13257,13742. .13804,13936. .14107,
                14566. .14833))
                /gene="At2g43320"
CDS             complement(join(11780. .11950,12040. .12519,12610. .12654,
                13002. .13102,13191. .13257,13742. .13804,13936. .14064))
                /gene="At2g43320"
                /codon_start=1
                /product="expressed protein"
                /protein_id="AAB64311.2"
                /db_xref="GI:20196858"
                /translation="MRAPSLLAQCLPGLLPQDRGGVSALSEKDLQLPTPAVEIIPSKT
                VAHHRYSGENLDALGLQVFKGKVSVADIIGLSGSETAPLKNEGSLKSWESSVVLVNVL
                KNEIRDGQLSFRGKRVLELGCNFGVPGIFACLKGASSVHFQDLSAETIRCTTIPNVLA
                NLEQARDRQSRQPESPLTPSRQAISASVRFYAGEWEELSTVLSIIRTDVLEPSIPAMN
                LSFSEEDFMDGCSSQDGSITGQPDFSSRRSRKLSGSRAWERANETGQEGEYGYDVILM
                TEIPYSITSLKKLYSLIKKCLRPPYGVMYLAAKKQYVGFNSGAKHLRNLVDEETILGA
                HLVKETTDRDIWKFPLK"
gene            16041. .18948
                /gene="At2g43330"
                /note="synonym: T1O24.7"
mRNA            join(<16041. .16188,16408. .16514,16563. .16898,17700. .17993,
                18080. .18948)
                /gene="At2g43330"
CDS             join(16041. .16188,16408. .16514,16563. .16898,17700. .17993,
                18080. .18760)
                /gene="At2g43330"
                /codon_start=1
                /product="putative membrane transporter"
                /protein_id="AAB64332.1"
                /db_xref="GI:2289003"
                /translation="MTLTIPNAPGSSGYLDMFPERRMSYFGNSYILGLTVTAGIGGLL
                FGYDTGVISGALLYIKDDFEVVKQSSFLQVYNVSSFTSSKLETIVSMALVGAMIGAAA
                GGWINDYYGRKKATLFADVVFAAGAIVMAAAPDPYVLISGRLLVGLGVGVASVTAPVY
                IAEASPSEVRGGLVSTNVLMITGGQFLSYLVNSAFTQVPGTWRWMLGVSGVPAVIQFI
                LMLFMPESPRWLFMKNRKAEAIQVLARTYDISRLEDEIDHLSAAEEEEKQRKRTVGYL
                DVFRSKELRLAFLAGAGLQAFQQFTGINTVMYYSPTIVQMAGFHSNQLALFLSLIVAA
                MNAAGTVVGIYFIDHCGRKKLALSSLFGVIISLLILSVSFFKQSETSSDGGLYGWLAV
                LGLALYIVFFAPGMGPVPWTVNSEIYPQQYRGICGGMSATVNWISNLIVAQTFLTIAE
                AAGTGMTFLILAGIAVLAVIFVIVFVPETQGLTFSEVEQIWKERAYGNISGWGSSSDS
                NNMEGLLEQGSQS"
repeat_region   complement(21849. .21900)
                /rpt_family="AT_rich"
gene            22675. .23510
                /gene="At2g43340"
                /note="synonym: T1O24.8"
mRNA            join(<22675. .23164,23243. .23304,23325. .23510)
                /gene="At2g43340"
CDS             join(22675. .23164,23243. .23304,23325. .23498)
                /gene="At2g43340"
                /codon_start=1
                /product="unknown protein"
                /protein_id="AAB64331.1"
                /db_xref="GI:2289002"
                /translation="MCDSDFLSRCSSYEADVDSDSEISTSSFSSCSYSESEEEEINNG
                FGVVQSKKQTKKLEKKKSNVLLEGYVVDSAVNDDLKRTKSLTDDDLEELKGCVDLGFG
                FNYEEIPELCNTLPALELCYSMSQKFIDQDHHHHSSSSPEKKSSVLDSPVSPIASWKI
                SSPGDNPDDVKARLKFWAQAVACTMRFGFDQPKAEESMEEEEDGETRQFAGVKSLMRL
                TCLKTTIVVELFAHTVCIKRLMI"
gene            complement(23549. .25505)
                /gene="At2g43350"
                /note="synonym: T1O24.9; supported by full length cDNA:
                Ceres:30737"
mRNA            complement(join(23549. .23749,24086. .24133,24215. .24382,
                24465. .24583,24673. .24734,24830. .24906,25278. .>25505))
                /gene="At2g43350"
CDS             complement(join(24101. .24133,24215. .24382,24465. .24583,
                24673. .24734,24830. .24906,25278. .25439))
                /gene="At2g43350"
                /codon_start=1
                /product="putative glutathione peroxidase"
                /protein_id="AAB64335.1"
                /db_xref="GI:2289006"
                /translation="MPRSSRWVNQRATSKIKKFILFLGVAFVFYLYRYPSSPSTVEQS
                STSIYNISVKDIEGKDVSLSKFTGKVLLIVNVASKCGLTHGNYKEMNILYAKYKTQGF
                EILAFPCNQFGSQEPGSNMEIKETVCNIFKAEFPIFDKIEVNGKNTCPLYNFLKEQKG
                GLFGDAIKWNFAKFLVDRQGNVVDRYAPTTSPLEIEKDIVKLLASA"
gene            complement(25700. .28116)
                /gene="At2g43360"
                /note="synonym: T1O24.10; supported by full length cDNA:
                Ceres:42038"
```

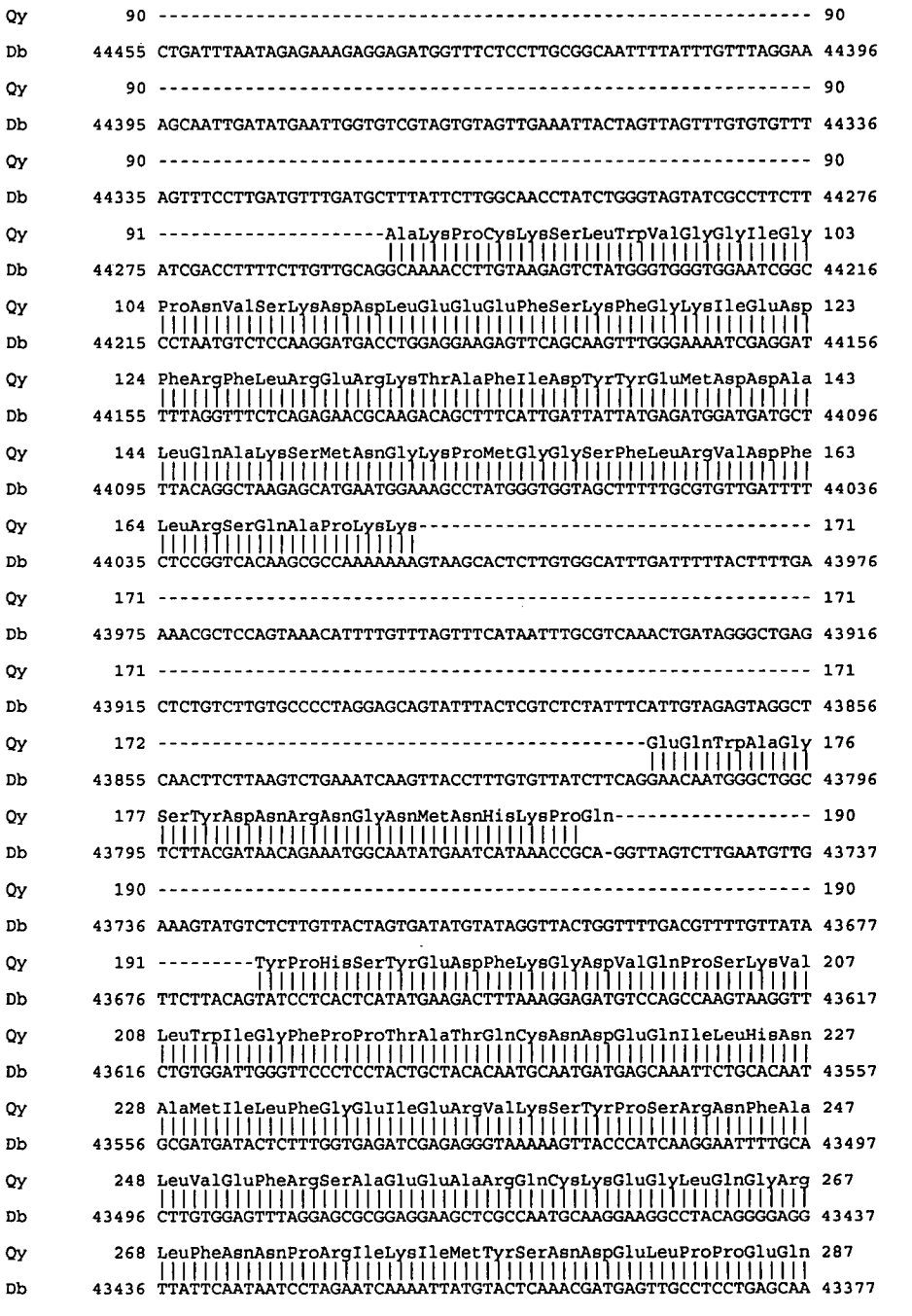
Alignment Scores:

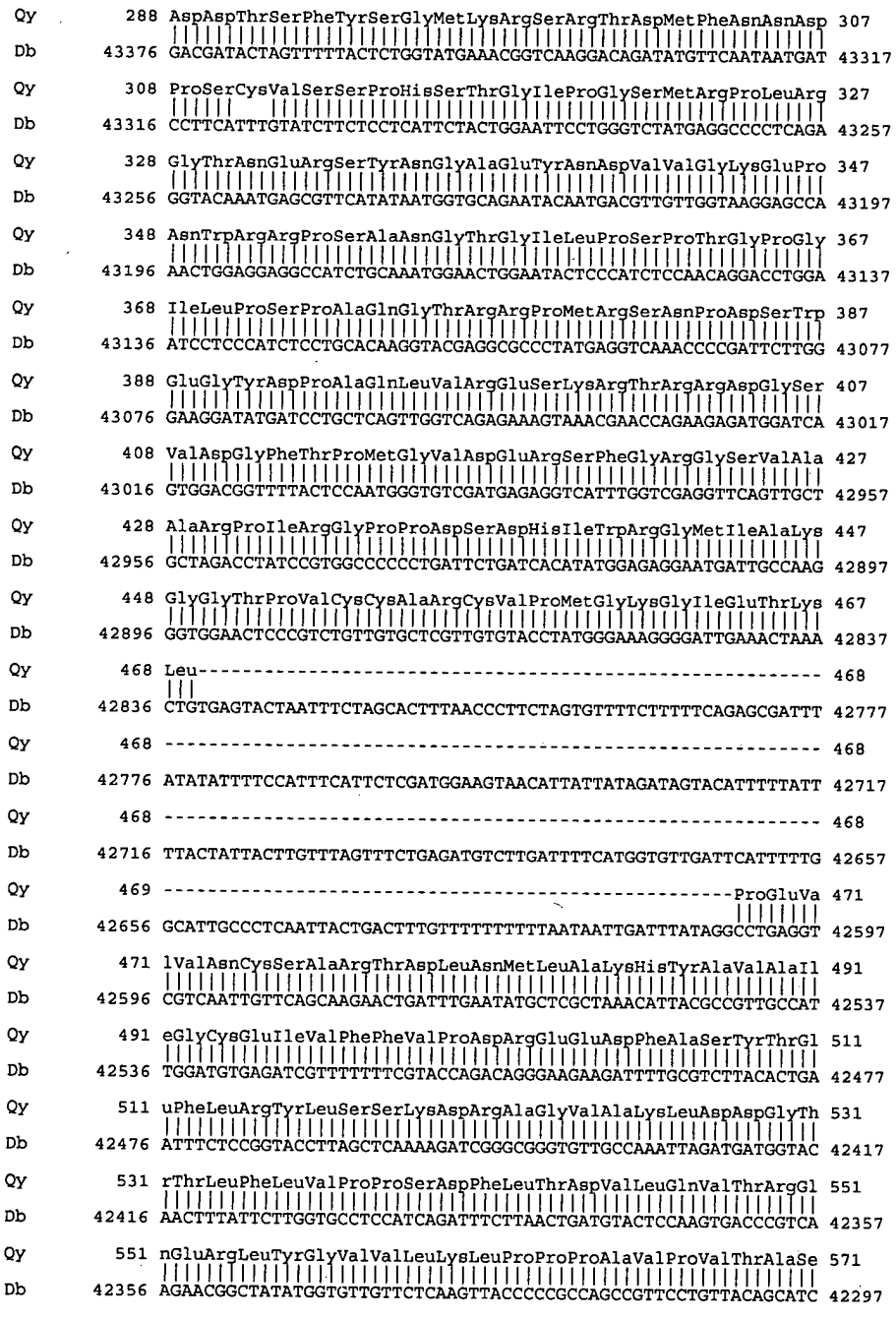
Pred. No.:	1.03e-165	Length:	102057
Score:	4392.00	Matches:	899
Percent Similarity:	58.72%	Conservative:	0
Best Local Similarity:	58.72%	Mismatches:	2
Query Match:	91.48%	Indels:	631
DB:	8	Gaps:	4

US-09-920-705-3 (1-901) x AC002335 (1-102057)

```
Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45895 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 45836

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45835 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 45776

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45775 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 45716

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45715 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 45656

Qy         81 SerGlnIleLysIleGluTyrAlaArgPro---------------------------- 90
              |||||||||||||||||||||||||||
Db      45655 AGTCAAATTAAGATCGAATACGCACGACCGGTTTGTTCTTATCTATATCTTCGTTTGTTC 45596

Qy         90 ------------------------------------------------------------ 90
```


Db 45595 TCTAACTTTGATTGTCTTTTGTCAACGATTATACTCTTTTTGCGAATTCATAGTCCAGGT 45536
Qy 90 ---------- 90
Db 45535 TCACAAAACTTTGATGATGCTTGTTTAGTCCAAAAATTTCTTGTTGAATCTGTTTTTTTTT 45476
Qy 90 ---------- 90
Db 45475 TTCCTCATCGTACAAATCAAAGTCGAAACCTAGTTTTTTTCTATTATACGTCGTTAGCTT 45416
Qy 90 ---------- 90
Db 45415 AAGGCGAAACCTGATCCGATCGAAACGTCTTTTCTCAAATTACTTTGGTTATATCGAACT 45356
Qy 90 ---------- 90
Db 45355 CGCGCAAAGCCAAACCACAGAGAAGCTCTGCAAAATTTGATGTTAAAGCATATATAACTC 45296
Qy 90 ---------- 90
Db 45295 TTAGCGAATGAGCTCTGCAAAAGAATACATCAAACACATGTTTACTCTCGTTTATGCGAA 45236
Qy 90 ---------- 90
Db 45235 GAAGGTTTAATCCGATTGTCGTTTATCTGCAAACTTGTATCCGCGTTTAGCAGTCTGATT 45176
Qy 90 ---------- 90
Db 45175 TCAGGTTCGCTTCAGATGTTAAATCTCACAAGCTTGAGTATGAATGTATTGCGTACTCCA 45116
Qy 90 ---------- 90
Db 45115 CTCGAAATTCGCAATGGTGAAAATTTGGAAGCGAGCAAACTCTATCATCGGCCAAACAGA 45056
Qy 90 ---------- 90
Db 45055 TAAGAAATTTTGGAGTTTAAAAGTTCAGTTTCCTGCAAAAATCAAACCGCGTGGAGAAAT 44996
Qy 90 ---------- 90
Db 44995 TTGTCTATGGCGGTACGTAGATATCAATAACTGCATCTGCGAAACAGAAAATTCTGGCAA 44936
Qy 90 ---------- 90
Db 44935 CCGTATCGTCCTAACTCCTAGTATCGCTGGCACATATCCATATCAGTGTGAGGGAGTTTT 44876
Qy 90 ---------- 90
Db 44875 GGGCTTGTTTTGGCTCGATGCTTCAGGAAGACAAATTACGTGGTTGTTAAGGCGGCTAAC 44816
Qy 90 ---------- 90
Db 44815 TCTACCAATCAGAAACGCTTTATTCGAAGAACCATGTTTGTTCCTCAATTCCCATCCCTA 44756
Qy 90 ---------- 90
Db 44755 CGTACAATCTGGGCTTTCCCATTGTAGTTCCTTAGGAAGTTGACTTCTTCACAACCATTC 44696
Qy 90 ---------- 90
Db 44695 TGGGATTGGTACAGTTGCAGTGGACAAGACAAATATTCATTTGCAGGCAGACTCAAATTT 44636
Qy 90 ---------- 90
Db 44635 CAATGTCTTGCCGGCTTTGAGTACTTAAATGGAGTGTTCAGGGATTGGTTTATTTGGGAC 44576
Qy 90 ---------- 90
Db 44575 TCAGGCAATGGATAAAGACAGGAATGTTTTGCAGAAAGTATTAATGTCTTTCCGGCTTTG 44516
Qy 90 ---------- 90
Db 44515 GGGACTACATACTTCTCCTACAGACAAAATCTGATGTTTAGGCAAGAGAAACTATTAACA 44456


Qy 90 ---------- 90
Db 44455 CTGATTTAATAGAGAAAGAGGAGATGGTTTCTCCTTGCGGCAATTTTATTTGTTTAGGAA 44396
Qy 90 ---------- 90
Db 44395 AGCAATTGATATGAATTGGTGTCGTAGTGTAGTTGAAATTACTAGTTAGTTTGTGTGTTT 44336
Qy 90 ---------- 90
Db 44335 AGTTTCCTTGATGTTTGATGCTTTATTCTTGGCAACCTATCTGGGTAGTATCGCCTTCTT 44276
Qy 91 ----------AlaLysProCysLysSerLeuTrpValGlyGlyIleGly 103
Db 44275 ATCGACCTTTTCTTGTTGCAGGCAAAACCTTGTAAGAGTCTATGGGTGGGTGGAATCGGC 44216
Qy 104 ProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAsp 123
Db 44215 CCTAATGTCTCCAAGGATGACCTGGAGGAAGAGTTCAGCAAGTTTGGGAAAATCGAGGAT 44156
Qy 124 PheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAla 143
Db 44155 TTTAGGTTTCTCAGAGAACGCAAGACAGCTTTCATTGATTATTATGAGATGGATGATGCT 44096
Qy 144 LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPhe 163
Db 44095 TTACAGGCTAAGAGCATGAATGGAAAGCCTATGGGTGGTAGCTTTTTGCGTGTTGATTTT 44036
Qy 164 LeuArgSerGlnAlaProLysLys---------- 171
Db 44035 CTCCGGTCACAAGCGCCAAAAAAAGTAAGCACTCTTGTGGCATTTGATTTTTACTTTTGA 43976
Qy 171 ---------- 171
Db 43975 AAACGCTCCAGTAAACATTTTGTTTAGTTTCATAATTTGCGTCAAACTGATAGGGCTGAG 43916
Qy 171 ---------- 171
Db 43915 CTCTGTCTTGTGCCCCTAGGAGCAGTATTTACTCGTCTCTATTTCATTGTAGAGTAGGCT 43856
Qy 172 ----------GluGlnTrpAlaGly 176
Db 43855 CAACTTCTTAAGTCTGAAATCAAGTTACCTTTGTGTTATCTTCAGGAACAATGGGCTGGC 43796
Qy 177 SerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGln---------- 190
Db 43795 TCTTACGATAACAGAAATGGCAATATGAATCATAAACCGCA-GGTTAGTCTTGAATGTTG 43737
Qy 190 ---------- 190
Db 43736 AAAGTATGTCTCTTGTTACTAGTGATATGTATAGGTTACTGGTTTTGACGTTTTGTTATA 43677
Qy 191 ---------TyrProHisSerTyrGluAspPheLysGlyAspValGlnProSerLysVal 207
Db 43676 TTCTTACAGTATCCTCACTCATATGAAGACTTTAAAGGAGATGTCCAGCCAAGTAAGGTT 43617
Qy 208 LeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHisAsn 227
Db 43616 CTGTGGATTGGGTTCCCTCCTACTGCTACACAATGCAATGATGAGCAAATTCTGCACAAT 43557
Qy 228 AlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsnPheAla 247
Db 43556 GCGATGATACTCTTTGGTGAGATCGAGAGGGTAAAAAGTTACCCATCAAGGAATTTTGCA 43497
Qy 248 LeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeuGlnGlyArg 267
Db 43496 CTTGTGGAGTTTAGGAGCGCGGAGGAAGCTCGCCAATGCAAGGAAGGCCTACAGGGGAGG 43437
Qy 268 LeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProProGluGln 287
Db 43436 TTATTCAATAATCCTAGAATCAAAATTATGTACTCAAACGATGAGTTGCCTCCTGAGCAA 43377



RESULT 7
AX420422/c
LOCUS AX420422 801 bp DNA linear PAT 18-JUN-2002
DEFINITION Sequence 4 from Patent WO0212518.
ACCESSION AX420422
VERSION AX420422.1 GI:21524578

```
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Amasino,R.M., Schomburg,F.M., Michaels,S.D. and Patton,D.
  TITLE     Floral induction gene
  JOURNAL   Patent: WO 0212518-A 4 14-FEB-2002;
            WISCONSIN ALUMNI RESEARCH FOUNDATION (US)
FEATURES             Location/Qualifiers
     source          1. .801
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
                     /note="FPA antisense fragment"
ORIGIN

Alignment Scores:
Pred. No.:                1.35e-49       Length:          801
Score:                    1436.00        Matches:         265
Percent Similarity:       99.62%         Conservative:    0
Best Local Similarity:    99.62%         Mismatches:      1
Query Match:              29.91%         Indels:          0
DB:                       6              Gaps:            0

US-09-920-705-3 (1-901) x AX420422 (1-801)

Qy        197 AspPheLysGlyAspValGlnProSerLysValLeuTrpIleGlyPheProProThrAla 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        799 GACTTTAAAGGAGATGTCCAGCCAAGTAAGGTTCTGTGGATTGGGTTCCCTCCTACTGCT 740

Qy        217 ThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGlu 236
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 ACACAATGCAATGATGAGCAAATTCTGCACAATGCGATGATACTCTTTGGTGAGATCGAG 680

Qy        237 ArgValLysSerTyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGlu 256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 AGGGTAAAAAGTTACCCATCAAGGAATTTTGCACTTGTGGAGTTTAGGAGCGCGGAGGAA 620

Qy        257 AlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIle 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 GCTCGCCAATGCAAGGAAGGCCTACAGGGGAGGTTATTCAATAATCCTAGAATCAAAATT 560

Qy        277 MetTyrSerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMet 296
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 ATGTACTCAAACGATGAGTTGCCTCCTGAGCAAGACGATACTAGTTTTTACTCTGGTATG 500

Qy        297 LysArgSerArgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSer 316
              ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db        499 AAACGGTCAAGGACAGATATGTTCAATAATGATCCTTCATTTGTATCTTCTCCTCATTCT 440

Qy        317 ThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGly 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 ACTGGAATTCCTGGGTCTATGAGGCCCCTCAGAGGTACAAATGAGCGTTCATATAATGGT 380

Qy        337 AlaGluTyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGly 356
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 GCAGAATACAATGACGTTGTTGGTAAGGAGCCAAACTGGAGGAGGCCATCTGCAAATGGA 320

Qy        357 ThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThr 376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 ACTGGAATACTCCCATCTCCAACAGGACCTGGAATCCTCCCATCTCCTGCACAAGGTACG 260

Qy        377 ArgArgProMetArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuVal 396
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 AGGCGCCCTATGAGGTCAAACCCCGATTCTTGGGAAGGATATGATCCTGCTCAGTTGGTC 200

Qy        397 ArgGluSerLysArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyVal 416
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 AGAGAAAGTAAACGAACCAGAAGAGATGGATCAGTGGACGGTTTTACTCCAATGGGTGTC 140

Qy        417 AspGluArgSerPheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAsp 436
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 GATGAGAGGTCATTTGGTCGAGGTTCAGTTGCTGCTAGACCTATCCGTGGCCCCCCTGAT 80

Qy        437 SerAspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrProValCysCysAlaArg 456
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 TCTGATCACATATGGAGAGGAATGATTGCCAAGGGTGGAACTCCCGTCTGTTGTGCTCGT 20

Qy        457 CysValProMetGlyLys 462
              |||||||||||||||||
Db         19 TGTGTACCTATGGGAAAG 2

RESULT 8
AK065222
LOCUS       AK065222               3543 bp    mRNA    linear   PLN 24-JUL-2003
DEFINITION  Oryza sativa (japonica cultivar-group) cDNA clone:J013002G20, full
            insert sequence.
ACCESSION   AK065222
VERSION     AK065222.1  GI:32975240
KEYWORDS    FLI_CDNA; CAP trapper.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   The Rice Full-Length cDNA Consortium, National Institute of
            Agrobiological Sciences Rice Full-Length cDNA Project Team:,
            Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
            Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
            Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
            Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
            Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K.,
            Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
            Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
            Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
            Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
            Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:,
            Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
            Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
            Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
            Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Yoshino,M. and Hayashizaki,Y.
  TITLE     Collection, mapping, and annotation of over 28,000 cDNA clones from
            japonica rice
  JOURNAL   Science 301 (5631), 376-379 (2003)
  MEDLINE   22752273
   PUBMED   12869764
REFERENCE   2  (bases 1 to 3543)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
            Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
            Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
            Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
            Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
            Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
            Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
            Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
            Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,
            Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K.,
            Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,
            Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K.,
            Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,
            Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S.,
            Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,
            Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A.,
            Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,
            Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and
            Yoshimura,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
            Agrobiological Sciences, Department of Molecular Genetics, Head of
```

Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT This clone is one of the 28K full-length cDNA clones from japonica rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.

FEATURES Location/Qualifiers
source 1. .3543
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J013002G20"

ORIGIN

Alignment Scores:

Pred. No.:	4.58e-46	Length:	3543
Score:	1365.00	Matches:	348
Percent Similarity:	52.59%	Conservative:	150
Best Local Similarity:	36.75%	Mismatches:	303
Query Match:	28.43%	Indels:	146
DB:	8	Gaps:	36

US-09-920-705-3 (1-901) x AK065222 (1-3543)

```
Qy         16 GlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThrGluSerAspLeuThr 35
Db        245 GAGACGAACACGCTGTGGGTGGGCAACCTGCCCGCGCAGGCGGCGGAGGACGACGTGATG 304

Qy         36 GluLeuPheGlyArgTyrGlyAspIleAspArgIleThrValTyrSerSer---ArgGly 54
Db        305 GCGTCCTTCTCCCCGCACGGCGCGCTCGACTGCGTCATGGCCCGCGCGGGCCCGCGGAGC 364

Qy         55 PheAlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGln 74
Db        365 TACGCGTTCGTGCTGTTCCGCTCCGTCCCGGAGGCCCGCGCCGCCCTCGACGCGCTCCAG 424

Qy         75 GlyAlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAlaArgProAlaLysProCys 94
Db        425 GGCTCCAAGGTCAAGGGCAGCGTCGTCCGGCTCGAGTTCGCCCGGCCGGCTAGAGCTGTT 484

Qy         95 LysSerLeuTrpValGlyGlyIleGlyProAsnValSerLysAspAspLeuGluGluGlu 114
Db        485 AAGAACTTATGGGTTGGTGGCATTAGCTCCTCAATTTCAAAGGAGGAGCTGGAGGAGGAG 544

Qy        115 PheSerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLysThrAlaPhe 134
Db        545 TTTAAGAAATTTGGAAAGGTTGATGGTATTGCATTCTCTCGTGATCAGACTTCCGCGTAC 604

Qy        135 IleAspTyrTyrGluMetAspAspAlaLeuGlnAla---LysSerMetAsnGlyLysPro 153
Db        605 ATTGATTTTGACAAGCTGGAAGATGCAATTTCGGCACATAGAGCCTTGAATGGAAGAGTT 664

Qy        154 MetGlyGlySerPheLeuArgValAspPheLeuArgSerGlnAlaProLysLysGluGln 173
Db        665 TTAGGTGGACAAGAACTGTGTGTTGATTTCCAAAGGTCCAGGGGC------AGAGCGGAA 718

Qy        174 Trp------AlaGlySerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGlnTyr 191
Db        719 TGGTTGGAGACAGGCAGCTTCAATGGTAGAACTGGACCTGCAAAG--------------- 763

Qy        192 ProHisSerTyrGluAspPheLysGlyAspValGlnProSerLysValLeuTrpIleGly 211
Db        764 ------GGTTATGGAGTACGAAATCGAGAATCCAATCCCACCAATGTTCTTTGGGTAGGT 817

Qy        212 PheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeu 231
Db        818 TTCCCTAATACAGCTAAA------ATCAATGAGGAGGCACTTCGGCAAGCGATGGCTGTA 871

Qy        232 PheGlyGluIleGluArgValLysSerTyrProSerArgAsnPheAlaLeuValGluPhe 251
Db        872 CATGGTGCTGTTACAAATACAAAAGTTTTCCCAACAAGGCAATATGCTTTTGTGGAGTTT 931

Qy        252 ArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsn 271
Db        932 GCAACAGTTGGGGAAGCTTCTAATGCTAAGAAAAATCTAGATGGTCGTCTTTTCAATGAT 991

Qy        272 ProArgIleLysIleMetTyrSerAsnAspGluLeuProProGluGln---AspAspThr 290
Db        992 CAGAGAATTCAAATTCTTTTCTCGAACAGTGAGCTTGCGCCAAATAAATTAGATAATCCA 1051

Qy        291 SerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsnAspProSerCys 310
Db       1052 ACAGCAGTTTCTGGATTTCCAAAATCA------GAGATGTATTATGATGATGGTCAGTAT 1105

Qy        311 ValSerSerProHisSerThrGlyIleProGlySerMetArgProLeuArgGlyThrAsn 330
Db       1106 GGCGCTTCTGATTAT--------------------TTTGATCCTCGACGGGGAAGATCA 1144

Qy        331 GluArgSerTyrAsnGlyAlaGluTyrAsnAspValValGlyLysGluProAsnTrpArg 350
Db       1145 AGATAC------------TTTGAGTACAGTGGTGTC---------------------- 1168

Qy        351 ArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuPro 370
Db       1169 ----------------------------------CCTGTTTCTGGTGGTATTCTTCCA 1192

Qy        371 SerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrpGluGlyTyr 390
Db       1193 TCACCGGAATCTGGA------AATCCTCTCTTAACTGGACGATCTGCTCAGAGTACATTC 1246

Qy        391 AspProAlaGlnLeuValArgGluSerLysArgThrArgArgAspGlySerValAsp--- 409
Db       1247 GATCCA------------AGAGAAGCCAAGAGGTTGCGGTTGGATGCTGCAGCTGACCCT 1294

Qy        410 ----------------------GlyPheThrProMetGlyValAspGluArgSerPhe 421
Db       1295 TATGATACAAGGGCAGGTAGTGAGGGTCTTTATTCTGCTGGGTACTCTCAGCGCGAGAGT 1354

Qy        422 GlyArgGlySerValAlaAlaArgPro---------IleArgGlyProProAspSer--- 437
Db       1355 GCACGTTCTGAGAGGAGCTCAAGTCCTGCTATTCGAATCCATGGCACAGTGCACCGAACA 1414

Qy        438 ---------AspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrProValCysCys 454
Db       1415 TCATACCTTGAGCACTTCTGGCGTGGTAGCATTGCCAAAGGTGGATCTCCTGTATGTCGT 1474

Qy        455 AlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluValValAsnCys 474
Db       1475 GCTCGCTGTTTGCCTATAAGGAAGGGTGTTGAGATACCTTTACCGGATGTTGTTAATTGT 1534
```

```
Qy       475 SerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGlu 494
Db      1535 TCCGCTAGAACTGGACTAGATATGCTGGCAAAGCATTATAGAGACGCTTCAGGGTTTGAT 1594

Qy       495 IleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArg 514
Db      1595 ATCGTCTTCTTCTTGCCAGATAGTGAAGATGACTTTGTTTCTTACACTGAATTTTTGCGC 1654

Qy       515 TyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPhe 534
Db      1655 TACTTAGGCTCAAAAAGTCGGGCAGGGGTTGTAAAGGTTGATGGGGGGACTACTTTATTT 1714

Qy       535 LeuValProProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeu 554
Db      1715 TTGGTCCCACCTTCCGATTTTTTACGAAATGTTTTGCAAGTTGATGGTCCAGAACGTCTT 1774

Qy       555 TyrGlyValValLeuLysLeuPro----------ProProAlaValProVal 568
Db      1775 TACGGTGTAGTATTACACATTCCGCAAATGTCTGCTGCTGCTCCTGCTTCTGCTCCCACT 1834

Qy       569 ThrAlaSerTyrArgGlnGlu---------SerGlnSerAsnProLeuHisTyrMetAsp 585
Db      1835 CCTGCTGTACAAAGGCCACAACTAACTGCACCAGAATCACAACCT------TACTACGAT 1888

Qy       586 GlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGluAsnTyr 605
Db      1889 GAA---AGGGAAATTCCTTTGCAGAGAAGGTATAGTATGATTACGCCTAGCAACAATCAC 1945

Qy       606 IleArgGlyAlaProGluHisLeuThrAla---------------------------- 615
Db      1946 CATCGAGATGCT---GATCACCGTGGATCTTTACGTGAGGATTCATTGCATCAATTGGGG 2002

Qy       616 ------AlaSerLysProSerValSerGluPro--------------------LeuArg 626
Db      2003 CAAATTCTTGCACGACCTCGGGTGGATGAAGGACAAGTAGTTCAGCCAAACCTTGCCGGT 2062

Qy       627 IleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAlaThr 646
Db      2063 ATCCCTACAAACGCAGGACTACAGGTACAGCCTTCACTCCAGCCTGATATGATAGCCACT 2122

Qy       647 LeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMet 666
Db      2123 TTAGCTAAACTTTGCCAAGTGGGCAGTCATCAGCTCTGGTCACTGGCCAGCTTCCCCTG 2182

Qy       667 SerGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyGlu 686
Db      2183 AGC------TCAACAGACCAACCTGCATTAACCCAAATGAAT---------GATGCATCT 2227

Qy       687 AlaProSerGlnAlaTrpLysArgGlyProGlnThrValHisAspAlaSerAsnGlnSer 706
Db      2228 ACTCTTGCAAAAGTATGGAGA------CCTGAAAATCAAGCAATGGCTTCTACTTCATCC 2281

Qy       707 PheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeuProProProProSerArg 726
Db      2282 TTAGAGCAGATCGGTAAT---TTCCAGCATTCAGGACAA---------CAGTTCAGCAAA 2329

Qy       727 TyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMetGln 746
Db      2330 CAAGCTGGAGCAGTTCATCTGCCAAACTAT------GGGAACTTGGCAGGTGCACAGGAG 2383

Qy       747 TyrGlnSerGln----------------------SerValAsnMetProGlnLeuSer 758
Db      2384 CACCCAACACAACATAGTGCTTACAATCCTGAAATGACTCTAAACTTGCCACCACCACCA 2443

Qy       759 ProLeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHis 778
Db      2444 CCACCTCCTACGCTGCCACCCAGTTCTGCTATATTATCATCGCAAGTTGGTCACAGTTTG 2503

Qy       779 ProValSerGlnProMetValGlnGlnTyrGlnProGluAla---SerMetProAsnGln 797
Db      2504 CCT---ACGCAGATGAGTCAGCAACAGTATCAACCAGAGCAGTATTATATGACTCAAAGC 2560

Qy       798 AsnTyrGlyProIleProSerTyrGlnGlnAlaAsnPheHisGly--------------- 812
Db      2561 AATTATGGTCAATTGGCCACAGTTAGCAGTTCTAATCTTCAGGCGCACCATCAACAAATT 2620

Qy       813 ValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGlnPro 832
Db      2621 GTTGCGACTCCTGCTGCACAAGCTCCTGTTGCTGCGCAATTCCCTCCTGCTATGCAAGCT 2680

Qy       833 ProAlaAspLysAlaAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMetIle 852
Db      2681 CCTGCT----------------------------------GCCGCACAAGCTCCTGTT 2704

Qy       853 SerGlyAspGlyGlnGlyThrThrAspGlyGluValAspLysAsnGlnArgTyrGlnSer 872
Db      2705 GCT------GCACAAGCTTCTGCTGAC---GAGGCAGAACGGAACAGGAAGTATCAGGCG 2755

Qy       873 ThrLeuGlnPheAlaAlaAsnLeuLeuLeuGlnIleGlnGlnLysGlnGlnGlnGlnSer 892
Db      2756 ACTCTTCAGTTGGCTCAGCGCCTGTTGGGTCAGCTACAACAGAAGCCTGGAAATCAACCT 2815

Qy       893 SerGlyThrProAlaGlyGln 899
Db      2816 TAAGGCTATCCTGGAAATCAA 2836
```

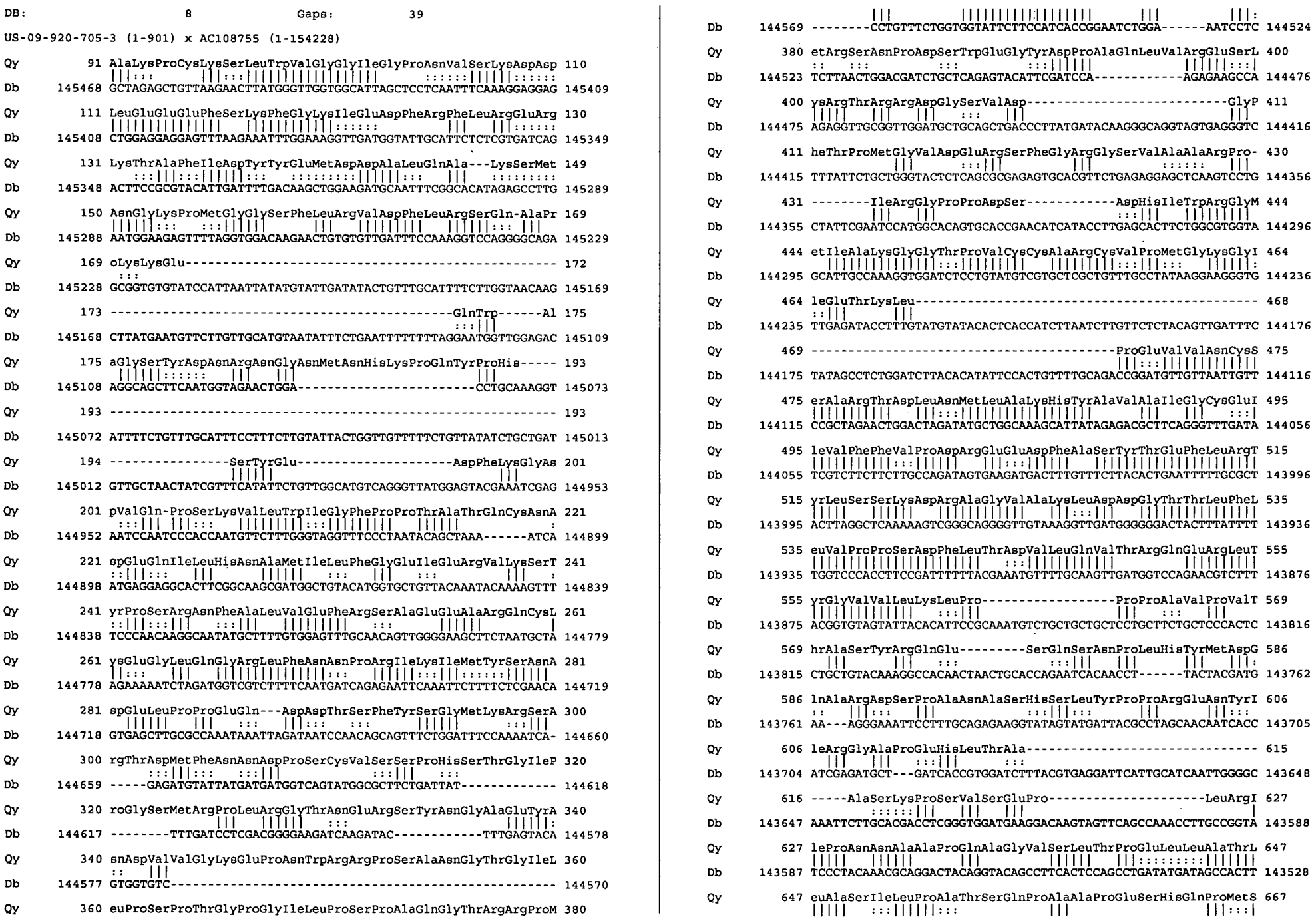
RESULT 9
AP005735/c
LOCUS AP005735 124083 bp DNA linear PLN 28-JUL-2004
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 9, BAC clone:OSJNBb0034B12.
ACCESSION AP005735
VERSION AP005735.3 GI:50726236
KEYWORDS .
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 9, BAC clone:OSJNBb0034B12
JOURNAL Published Only in Database (2002)
REFERENCE 2 (bases 1 to 124083)
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Direct Submission
JOURNAL Submitted (11-SEP-2002) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT On Jul 27, 2004 this sequence version replaced gi:40736955.
Genes were predicted from the integrated results of the following: GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH (http://www.softberry.com/), GeneMark.hmm (http://opal.biology.gatech.edu/GeneMark/), GlimmerM (http://www.tigr.org/tdb/glimmerm/glmr_form.html), RiceHMM (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4 (http://globin.cse.psu.edu/html/docs/sim4.html), gap2 (http://www.tigr.org/software/glimmerm/), BLASTN and BLASTX. The genomic sequence was searched against NCBI NonRedundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP or DDBJ. Protein homologies of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no.
A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted

```
by two or more gene prediction programs is classified as a
'hypothetical' protein according to IRGSP standard. A gene
predicted by a single gene prediction program is also classified as
a probable 'hypothetical' protein and is included as a
miscellaneous feature of the sequence.
The orientation of the sequence is from M13rev to -21M13 of the BAC
clone. This sequence of OSJNBb0034B12 has an overlap with
P0450E05(DDBJ: AP006756) clone at 5' end and with P0217C03(DDBJ:
AP006173) clone at 3' end. Detailed information on overlap and
assembly quality together with annotation of this entry is
available at http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             Location/Qualifiers
     source          1. .124083
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="9"
                     /clone="OSJNBb0034B12"
     gene            3185. .6071
                     /gene="OSJNBb0034B12.1"
     mRNA            join(3185. .3279,3941. .4032,4234. .4283,4368. .4450,
                     4600. .4708,5392. .5493,5640. .6071)
                     /gene="OSJNBb0034B12.1"
                     /note="supported by full-length cDNA(s): AK103734"
     CDS             join(3241. .3279,3941. .4032,4234. .4283,4368. .4450,
                     4600. .4708,5392. .5493,5640. .5686)
                     /gene="OSJNBb0034B12.1"
                     /note="contains full-length cDNA(s): AK103734"
                     /codon_start=1
                     /product="unknown protein"
                     /protein_id="BAD33813.1"
                     /db_xref="GI:50726237"
                     /translation="MAMAMGLLAKNPKMINHRYASDMQLQHRLSPACSVMFNKQCSYR
                     ITRKACSVLGAVSPIQCTETSTESLVSFKDFLVSVQTEEDGLIKLRVTVADTMTESIF
                     EKVFSKNVAAAQPLPGFRRMKGDTKRDCSALNWAIKGEERNHQEYHQPHNCRICPKGG
                     PRCIKKSEGSTDL"
     gene            7566. .12011
                     /gene="OSJNBb0034B12.2"
     mRNA            join(7566. .7812,7988. .8078,8413. .8491,8850. .9183,
                     9374. .10408,10581. .10730,11221. .12011)
                     /gene="OSJNBb0034B12.2"
                     /note="supported by full-length cDNA(s): AK065778"
     CDS             join(7786. .7812,7988. .8078,8413. .8491,8850. .9183,
                     9374. .10408,10581. .10730,11221. .11655)
                     /gene="OSJNBb0034B12.2"
                     /note="contains EST(s): C98201(C1066),D15674(C1066A)
                     contains full-length cDNA(s): AK065778"
                     /codon_start=1
                     /product="putative translation initiation factor IF-2"
                     /protein_id="BAD33814.1"
                     /db_xref="GI:50726238"
                     /translation="MAWRMLRRKDFHTGLVNLASRVDHGGAKNFSSGTFGKLADFVLS
                     DTHTPIVKGAANCTAYKHCTIRNFHAGVYMLAWSRKREEVAGLKAPKKEKRVKRETRT
                     QPPVEAPYVAPKQKIAIKSSPDKTVDIFDGMTLLDLSKRTGASIGALQDILTDLGEKV
                     ESEFDAISIDLAELVAMELGVNIRRMHTGEGTLEPRPAVVTIMGHVDHGKTSLLDSLR
                     QTSVAAKEAGGITQHIGAFVVEMPSGASITFLDTPGHAAFSAMRARGAAVTDIVVLVV
                     AADDGVMPQTLEAMSHAKAANVPIVVAVNKCDKSGADPERVRIQLGSEGLLLEDMGGD
                     VQVVEISAVTKLGLDKLEEALLLQAEIMDLKARIDGPAQAFVVEARVDRGRGPLATAI
                     VKAGTLVSGQHIVVGAEWGRIRSLRDTAGKITESAKPAMPVEIEGLRGLPMAGDDVVV
                     VDSEERARMLSQGRKKKQEKDRLRKIDEDMTEEAEIGEETPERVEMPIIVKADVQGSV
                     QAVTDALRSLNSPQVFVNIVHVGVGPISQHDIDLAQACRAYIVGFNIRTPPSAITLAA
                     TQANIKILLHKVIYHLLEEMGREIVEKAPGTPETQVSGEAEVLNIFELKGRSKSKGPD
                     IKIAGCRITDGHLSKTGTMRLLRSGDVVFEGPCASLKREKQDAETVDKGNDCGLVIQD
                     CNDFQVGDIVQCLEQVIRKPKFISTQSGAVRIEC"
     gene            complement(12465. .12713)
                     /gene="OSJNBb0034B12.3"
     mRNA            complement(<12465. .>12713)
                     /gene="OSJNBb0034B12.3"
                     /note="start and end point are not identified"
     CDS             complement(12465. .12713)
                     /gene="OSJNBb0034B12.3"
                     /note="predicted by FGENESH etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD33815.1"
                     /db_xref="GI:50726239"
                     /translation="MAGSKSVRVILLAVVLAVFAVLLAGAAAARPAPAAERSGGEAAA
                     AAYVAVNPAAAVVEKARETVEMLMARLPAGPSPKGPGH"
     gene            13577. .15039
                     /gene="OSJNBb0034B12.4"
     mRNA            join(<13577. .13620,13698. .13844,14101. .14263,
                     14944. .>15039)
                     /gene="OSJNBb0034B12.4"
                     /note="start and end point are not identified"
     CDS             join(13577. .13620,13698. .13844,14101. .14263,14944. .15039)
                     /gene="OSJNBb0034B12.4"
                     /note="predicted by GENSCAN etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD33816.1"
                     /db_xref="GI:50726240"
                     /translation="MMKWNLIPMPFVSNCVQDPRIESFIISNRQAAKPEERSYYVPKV
                     PRMVREIGNERNNEHFSSESLSDLMLALIQQREKKDRTDDLNVQVQDEIGMVREDSKV
                     RILAVEMILWHKRWHMLVLAEYQVKRTSSENDSGVQVQDEIRTATTR"
     gene            complement(15987. .19946)
                     /gene="OSJNBb0034B12.5"
     mRNA            complement(join(15987. .16346,16512. .16613,16686. .16771,
                     17095. .17366,17870. .17961,18055. .18192,18269. .18339,
                     18646. .18829,18907. .19039,19421. .19936))
                     /gene="OSJNBb0034B12.5"
                     /note="supported by full-length cDNA(s): AK066076"
     mRNA            complement(join(16003. .16346,16512. .16613,16686. .16771,
                     17095. .17366,17870. .17961,18055. .18192,18269. .18339,
                     18646. .18829,18907. .19039,19421. .19946))
                     /gene="OSJNBb0034B12.5"
                     /note="supported by full-length cDNA(s): AK070794"
     CDS             complement(join(16185. .16346,16512. .16613,16686. .16771,
                     17095. .17366,17870. .17961,18055. .18192,18269. .18339,
                     18646. .18829,18907. .19039,19421. .19515))
                     /gene="OSJNBb0034B12.5"
                     /note="contains EST(s): AU085809(C53716),C28019(C53716)
                     contains full-length cDNA(s): AK066076,AK070794"
                     /codon_start=1
                     /product="putative tbc1 domain family protein"
                     /protein_id="BAD33817.1"
                     /db_xref="GI:50726241"
                     /translation="MSGGGGGGGGGGGGSPNNTEWRFNQTLRNVQGMLKGRSFPGKVL
                     LTRRSEPLSPPEYSPRYENDRDEYEQNEGSQEGKGQASGNTADSMSAKKSNPPSTSST
                     NSLPDAQGLVSGARATDSARIAKFTNELSRPAVILDKLRELSWSGVPPYMRPNIWRLL
                     LGYAPPNADRREGVLTRKRLEYVECVSQYYDIPDTERSDEEINMLRQIAVDCPRTVPD
                     VTFFQHPQIQKSLERILYTWAIRHPASGYVQGINDLLTPFLVVFLSEHLEGNMDTWSM
                     EKLSPQDVSNIEADCYWCLSKFLDGMQDHYTFAQPGIQRLVFRLKELVHRIDEPVSKH
                     MEEQGLDFLQFAFRWFNCLMIREIPFHLVTRLWDTYLAEGDYLPDFLVYISASFLLTW
                     SDKLKKLDFQEMVMFLQHLPTRNWAHHELEMVLSRAYMWHTMFKSSPSHLAS"
     gene            20737. .21569
                     /gene="OSJNBb0034B12.6"
     mRNA            join(<20737. .20769,21306. .>21569)
                     /gene="OSJNBb0034B12.6"
                     /note="start and end point are not identified"
     CDS             join(20737. .20769,21306. .21569)
                     /gene="OSJNBb0034B12.6"
                     /note="predicted by FGENESH etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD33818.1"
                     /db_xref="GI:50726242"
                     /translation="MSVATNLQHITHAPRSNRLGLAGRASRRAWTENAEEECGGGVSF
                     GAAGEDGHGGDGDGDGVNAYLRFPGPTLTGSGDGGVVGGEVDGDATLLLMAGLI"
     gene            complement(22328. .22807)
                     /gene="OSJNBb0034B12.7"
     mRNA            complement(<22328. .>22807)
                     /gene="OSJNBb0034B12.7"
                     /note="start and end point are not identified"
     CDS             complement(22328. .22807)
                     /gene="OSJNBb0034B12.7"
```

```
                     /note="predicted by GeneMark.hmm etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD33819.1"
                     /db_xref="GI:50726243"
                     /translation="MPAPASSRYVPLAAAPEQRHDGEAEECCYAATTTRYVPLRRRRG

Alignment Scores:
Pred. No.:                1.63e-34      Length:          124083
Score:                    1117.50       Matches:         325
Percent Similarity:       45.76%        Conservative:    134
Best Local Similarity:    32.40%        Mismatches:      280
Query Match:              23.28%        Indels:          264
DB:                       8             Gaps:            39

US-09-920-705-3 (1-901) x AP005735 (1-124083)

Qy         91 AlaLysProCysLysSerLeuTrpValGlyGlyIleGlyProAsnValSerLysAspAsp 110
Db      42641 GCTAGAGCTGTTAAGAACTTATGGGTTGGTGGCATTAGCTCCTCAATTTCAAAGGAGGAG 42582

Qy        111 LeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArg 130
Db      42581 CTGGAGGAGGAGTTTAAGAAATTTGGAAAGGTTGATGGTATTGCATTCTCTCGTGATCAG 42522

Qy        131 LysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGlnAla---LysSerMet 149
Db      42521 ACTTCCGCGTACATTGATTTTGACAAGCTGGAAGATGCAATTTCGGCACATAGAGCCTTG 42462

Qy        150 AsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeuArgSerGln-AlaPr 169
Db      42461 AATGGAAGAGTTTTAGGTGGACAAGAACTGTGTGTTGATTTCCAAAGGTCCAGGGGCAGA 42402

Qy        169 oLysLysGlu------------------------------------------------- 172
Db      42401 GCGGTGTGTATCCATTAATTATATGTATTGATATACTGTTTGCATTTTCTTGGTAACAAG 42342

Qy        173 ----------------------------------------------GlnTrp------Al 175
Db      42341 CTTATGAATGTTCTTGTTGCATGTAATATTTCTGAATTTTTTTTAGGAATGGTTGGAGAC 42282

Qy        175 aGlySerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGlnTyrProHis----- 193
Db      42281 AGGCAGCTTCAATGGTAGAACTGGA----------------------CCTGCAAAGGT 42246

Qy        193 ------------------------------------------------------------ 193
Db      42245 ATTTTCTGTTTGCATTTCCTTTCTTGTATTACTGGTTGTTTTTCTGTTATATCTGCTGAT 42186

Qy        194 ----------------SerTyrGlu--------------------AspPheLysGlyAs 201
Db      42185 GTTGCTAACTATCGTTTCATATTCTGTTGGCATGTCAGGGTTATGGAGTACGAAATCGAG 42126

Qy        201 pValGln-ProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnA 221
Db      42125 AATCCAATCCCACCAATGTTCTTTGGGTAGGTTTCCCTAATACAGCTAAA------ATCA 42072

Qy        221 spGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerT 241
Db      42071 ATGAGGAGGCACTTCGGCAAGCGATGGCTGTACATGGTGCTGTTACAAATACAAAAGTTT 42012

Qy        241 yrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysL 261
Db      42011 TCCCAACAAGGCAATATGCTTTTGTGGAGTTTGCAACAGTTGGGGAAGCTTCTAATGCTA 41952

Qy        261 ysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsnA 281
Db      41951 AGAAAAATCTAGATGGTCGTCTTTTCAATGATCAGAGAATTCAAATTCTTTTCTCGAACA 41892

Qy        281 spGluLeuProProGluGln---AspAspThrSerPheTyrSerGlyMetLysArgSerA 300
Db      41891 GTGAGCTTGCGCCAAATAAATTAGATAATCCAACAGCAGTTTCTGGATTTCCAAAATCA- 41833

Qy        300 rgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIleP 320
Db      41832 -----GAGATGTATTATGATGATGGTCAGTATGGCGCTTCTGATTAT------------ 41791

Qy        320 roGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrA 340
Db      41790 --------TTTGATCCTCGACGGGGAAGATCAAGATAC------------TTTGAGTACA 41751

Qy        340 snAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleL 360
Db      41750 GTGGTGTC---------------------------------------------------- 41743

Qy        360 euProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgProM 380
Db      41742 --------CCTGTTTCTGGTGGTATTCTTCCATCACCGGAATCTGGA------AATCCTC 41697

Qy        380 etArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuValArgGluSerL 400
Db      41696 TCTTAACTGGACGATCTGCTCAGAGTACATTCGATCCA------------AGAGAAGCCA 41649

Qy        400 ysArgThrArgArgAspGlySerValAsp-------------------------GlyP 411
Db      41648 AGAGGTTGCGGTTGGATGCTGCAGCTGACCCTTATGATACAAGGGCAGGTAGTGAGGGTC 41589

Qy        411 heThrProMetGlyValAspGluArgSerPheGlyArgGlySerValAlaAlaArgPro- 430
Db      41588 TTTATTCTGCTGGGTACTCTCAGCGCGAGAGTGCACGTTCTGAGAGGAGCTCAAGTCCTG 41529

Qy        431 --------IleArgGlyProProAspSer------------AspHisIleTrpArgGlyM 444
Db      41528 CTATTCGAATCCATGGCACAGTGCACCGAACATCATACCTTGAGCACTTCTGGCGTGGTA 41469

Qy        444 etIleAlaLysGlyGlyThrProValCysCysAlaArgCysValProMetGlyLysGlyI 464
Db      41468 GCATTGCCAAAGGTGGATCTCCTGTATGTCGTGCTCGCTGTTTGCCTATAAGGAAGGGTG 41409

Qy        464 leGluThrLysLeu--------------------------------------------- 468
Db      41408 TTGAGATACCTTTGTATGTATACACTCACCATCTTAATCTTGTTCTCTACAGTTGATTTC 41349

Qy        469 ----------------------------------------ProGluValValAsnCysS 475
Db      41348 TATAGCCTCTGGATCTTACACATATTCCACTGTTTTGCAGACCGGATGTTGTTAATTGTT 41289

Qy        475 erAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGluI 495
Db      41288 CCGCTAGAACTGGACTAGATATGCTGGCAAAGCATTATAGAGACGCTTCAGGGTTTGATA 41229

Qy        495 leValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgT 515
Db      41228 TCGTCTTCTTCTTGCCAGATAGTGAAGATGACTTTGTTTCTTACACTGAATTTTTGCGCT 41169

Qy        515 yrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheL 535
Db      41168 ACTTAGGCTCAAAAAGTCGGGCAGGGGTTGTAAAGGTTGATGGGGGGACTACTTTATTTT 41109

Qy        535 euValProProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuT 555
Db      41108 TGGTCCCACCTTCCGATTTTTTACGAAATGTTTTGCAAGTTGATGGTCCAGAACGTCTTT 41049

Qy        555 yrGlyValValLeuLysLeuPro------------------ProProAlaValProValT 569
Db      41048 ACGGTGTAGTATTACACATTCCGCAAATGTCTGCTGCTGCTCCTGCTTCTGCTCCCACTC 40989

Qy        569 hrAlaSerTyrArgGlnGlu---------SerGlnSerAsnProLeuHisTyrMetAspG 586
Db      40988 CTGCTGTACAAAGGCCACAACTAACTGCACCAGAATCACAACCT------TACTACGATG 40935

Qy        586 lnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGluAsnTyrI 606
Db      40934 AA---AGGGAAATTCCTTTGCAGAGAAGGTATAGTATGATTACGCCTAGCAACAATCACC 40878

Qy        606 leArgGlyAlaProGluHisLeuThrAla------------------------------ 615
```

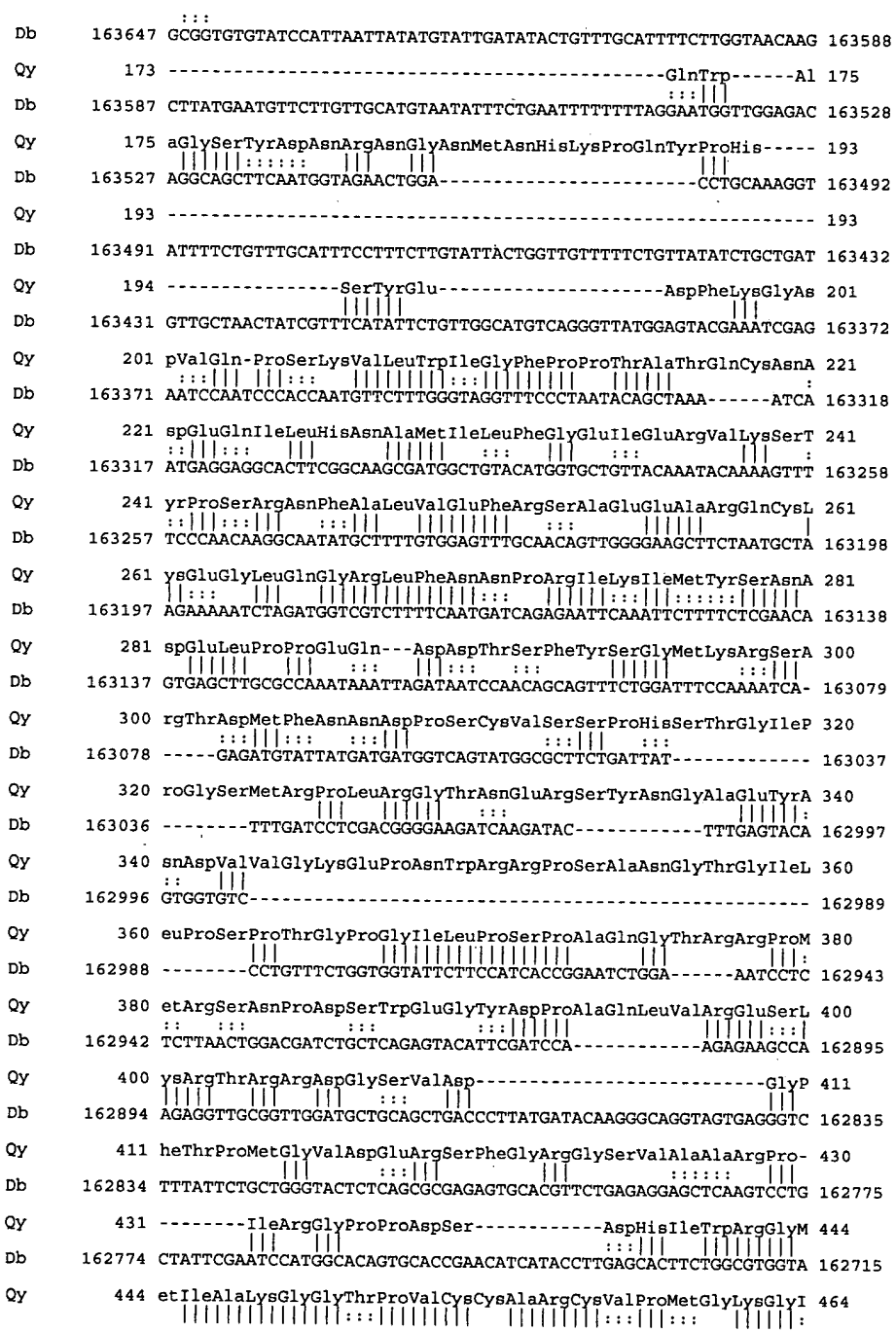
```
Db      40877 ATCGAGATGCT---GATCACCGTGGATCTTTACGTGAGGATTCATTGCATCAATTGGGGC 40821

Qy        616 -----AlaSerLysProSerValSerGluPro--------------------LeuArgI 627

Db      40820 AAATTCTTGCACGACCTCGGGTGGATGAAGGACAAGTAGTTCAGCCAAACCTTGCCGGTA 40761

Qy        627 leProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAlaThrL 647

Db      40760 TCCCTACAAACGCAGGACTACAGGTACAGCCTTCACTCCAGCCTGATATGATAGCCACTT 40701

Qy        647 euAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMetS 667

Db      40700 TAGCTAAACTTTTGCCAAGTGGGCAGTCATCAGCTCTGGTCACTGGCCAGCTTCCCCTGA 40641

Qy        667 erGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyGluA 687

Db      40640 GC------TCAACAGACCAACCTGCATTAACCCAAATGAAT---------GATGCATCTA 40596

Qy        687 laProSerGlnAlaTrpLysArgGlyProGlnThrValHisAspAlaSerAsnGlnSerP 707

Db      40595 CTCTTGCAAAAGTATGGAGA------CCTGAAAATCAAGCAATGGCTTCTACTTCATCCT 40542

Qy        707 heGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeuProProProProSerArgT 727

Db      40541 TAGAGCAGATCGGTAAT---TTCCAGCATTCAGGACAA---------CAGTTCAGCAAAC 40494

Qy        727 yrProProAlaSerAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMetGlnT 747

Db      40493 AAGCTGGAGCAGTTCATCTGCCAAACTAT------GGGAACTTGGCAGGTGCACAGGAGC 40440

Qy        747 yrGlnSerGln-----------------------SerValAsnMetProGlnLeuSerP 759

Db      40439 ACCCAACACAACATAGTGCTTACAATCCTGAAATGACTCTAAACTTGCCACCACCACCAC 40380

Qy        759 roLeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisP 779

Db      40379 CACCTCCTACGCTGCCACCCAGTTCTGCTATATTATCATCGCAAGTTGGACACAGTTTGC 40320

Qy        779 roValSerGlnProMetValGlnGlnTyrGlnProGluAla---SerMetProAsnGlnA 798

Db      40319 CT---ACGCAGATGAGTCAGCAACAGTATCAACCAGAGCAGTATTATATGACTCAAAGCA 40263

Qy        798 snTyrGlyProIleProSerTyrGlnGlnAlaAsn-PheHisGlyVal------------ 813

Db      40262 ATTATGGTCAATTGGCCACAGTTAGCAGTTCTAATCTTCAGGTCAGCAATTCTAACATCC 40203

Qy        814 --------------------------------ThrThrAsnGlnAlaGlnAsnLeuAsn 822

Db      40202 CATTTCTTCCTTATTCCCAAGTAAACCTTGCCAACAAATAATCAGATAGGGTTTTTG--- 40146

Qy        823 ProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsn------------ 838

Db      40145 ---GCAGTGTTTCAGCATGCTTTGACTTTGCCTAATGATAGGGTAAACTATGGTTTCTCT 40089

Qy        839 LeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAsp--------- 855

Db      40088 TTACAGGCGCACCATCAACAAATTGTTGCGACTCCTGCTGCACAAGCTCCTGTTGCTGCG 40029

Qy        856 ----------------------------------------------------GlyGlnGly 858

Db      40028 CAATTCCCTCCTGCTATGCAAGCTCCTGCTGCCGCACAAGCTCCTGTTGCTGCACAAGCT 39969

Qy        859 ThrThrAspGlyGluValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAla 878

Db      39968 TCTGCTGAC---GAGGCAGAACGGAACAGGAAGTATCAGGCGACTCTTCAGTTGGCTCAG 39912

Qy        879 AsnLeuLeuLeuGlnIleGlnGlnLysGlnGlnGlnGlnSerSerGlyThrProAlaGly 898

Db      39911 CGCCTGTTGGGTCAGCTACAACAGAAGCCTGGAAATCAACCTTAAGGCTATCCTGGAAAT 39852

Qy        899 Gln 899
```

```
Db      39851 CAA 39849

RESULT 10
AC108755/c
LOCUS       AC108755               154228 bp    DNA     linear   PLN 16-JUL-2004
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 9 BAC clone
            OSJNBa0028C22, complete sequence.
ACCESSION   AC108755
VERSION     AC108755.2  GI:50345123
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 154228)
  AUTHORS   Lee,J.-S., Hahn,J.-H., Yoon,U.-H., Lee,K.-S., Kim,Y.-H., Eun,M.Y.,
            Kim,H.-I., McCombie,W.R., Zutavern,T., de la Bastide,M.,
            Spiegel,L., Muller,S., Nascimento,L., Balija,V., Bell,M.,
            Miller,B., Katzenberger,F., Andrade,M.V., Dike,S., O'Shaughnessy,A.
            and Palmer,L.
  TITLE     Oryza sativa (japonica cultivar-group) chromosome 9 BAC clone
            OSJNBa0028C22, complete sequence
  JOURNAL   Unpublished (2004)
REFERENCE   2  (bases 1 to 154228)
  AUTHORS   Hahn,J.-H. and Kim,H.-I.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2002) Rice Genome Sequencing Project, National
            Institute of Agricultural Science and Technology(NIAST), RDA, 249
            Seodun-dong, Suwon 441-707, Korea (E-mail:jhhahn@rda.go.kr,
            Tel:82-31-290-0309, Fax:82-31-290-0308)
REFERENCE   3  (bases 1 to 154228)
  AUTHORS   Hahn,J.-H. and Kim,H.-I.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2004) Rice Genome Sequencing Project, National
            Institute of Agricultural Biotechnology (NIAB), RDA, 225
            Seodun-dong, Suwon 441-707, Korea
COMMENT     On Jul 16, 2004 this sequence version replaced gi:18449960.
            This sequence was finished as follows unless otherwise noted:  all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. OSJNBa0028C22
            overlaps clone OSJNBa0065A15 (AC137594) from base 1 to base 12997.
            The overlap is from base 155808 to base 168804 on OSJNBa0065A15.
            OSJNBa0028C22 also overlaps clone OSJNBb0034B12 (AP005735) from
            base 102828 to base 154228.  The overlap is from base 1 to base
            51401 on OSJNBb0034B12. This BAC clone was sequenced to phase II by
            the Korean National Institute of Agricultural Biotechnology and was
            completed to phase III by the Cold Spring Harbor Laboratory Genome
            Center.
FEATURES             Location/Qualifiers
     source          1. .154228
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="9"
                     /clone="OSJNBa0028C22"
                     /clone_lib="HindIII"
     misc_feature    56468. .56601
                     /note="The assembly is covered by high quality sequences
                     derived from a transposed plasmid subclone."
ORIGIN

Alignment Scores:
Pred. No.:              2.06e-34        Length:         154228
Score:                  1117.50         Matches:        325
Percent Similarity:     45.76%          Conservative:   134
Best Local Similarity:  32.40%          Mismatches:     280
Query Match:            23.28%          Indels:         264
```


DB: 8 Gaps: 39
US-09-920-705-3 (1-901) x AC108755 (1-154228)

```
Db     143527 TAGCTAAACTTTTGCCAAGTGGGCAGTCATCAGCTCTGGTCACTGGCCAGCTTCCCCTGA 143468

Qy        667 erGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyGluA 687
              ||     ||||||                 |||   |||         :::
Db     143467 GC------TCAACAGACCAACCTGCATTAACCCAAATGAAT---------GATGCATCTA 143423

Qy        687 laProSerGlnAlaTrpLysArgGlyProGlnThrValHisAspAlaSerAsnGlnSerP 707
                 ::::::   |||:::      |||:::              ||||||      |||
Db     143422 CTCTTGCAAAAGTATGGAGA------CCTGAAAATCAAGCAATGGCTTCTACTTCATCCT 143369

Qy        707 heGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeuProProProProSerArgT 727
               :::|||   ||||||   :::      :::|||||                 |||:::
Db     143368 TAGAGCAGATCGGTAAT---TTCCAGCATTCAGGACAA---------CAGTTCAGCAAAC 143321

Qy        727 yrProProAlaSerAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMetGlnT 747
                  |||   :::   |||||||||      |||   :::   |||      ::::
Db     143320 AAGCTGGAGCAGTTCATCTGCCAAACTAT------GGGAACTTGGCAGGTGCACAGGAGC 143267

Qy        747 yrGlnSerGln-----------------------SerValAsnMetProGlnLeuSerP 759
              ::   :::|||                     ::::::|||:::|||          |
Db     143266 ACCCAACACAACATAGTGCTTACAATCCTGAAATGACTCTAAACTTGCCACCACCACCAC 143207

Qy        759 roLeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisP 779
              ||   |||   :::|||   ::::::      :::   :::|||          :::   |
Db     143206 CACCTCCTACGCTGCCACCCAGTTCTGCTATATTATCATCGCAAGTTGGACACAGTTTGC 143147

Qy        779 roValSerGlnProMetValGlnGlnTyrGlnProGluAla---SerMetProAsnGlnA 798
              ||   :::|||           ||||||||||||||||      |||          |
Db     143146 CT---ACGCAGATGAGTCAGCAACAGTATCAACCAGAGCAGTATTATATGACTCAAAGCA 143090

Qy        798 snTyrGlyProIleProSerTyrGlnGlnAlaAsn-PheHisGlyVal------------ 813
              |||||||   :::   :::          :::||| |||
Db     143089 ATTATGGTCAATTGGCCACAGTTAGCAGTTCTAATCTTCAGGTCAGCAATTCTAACATCC 143030

Qy        814 --------------------------------ThrThrAsnGlnAlaGlnAsnLeuAsn 822
                                             |||   ||||||          |||
Db     143029 CATTTCTTCCTTATTCCCAAGTAAACCTTGCCAACAAATAATCAGATAGGGTTTTTG--- 142973

Qy        823 ProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsn------------ 838
                 :::   ||||||   |||:::      |||   |||:::   |||
Db     142972 ---GCAGTGTTTCAGCATGCTTTGACTTTGCCTAATGATAGGGTAAACTATGGTTTCTCT 142916

Qy        839 LeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAsp--------- 855
              |||:::      :::|||   :::          |||
Db     142915 TTACAGGCGCACCATCAACAAATTGTTGCGACTCCTGCTGCACAAGCTCCTGTTGCTGCG 142856

Qy        856 ----------------------------------------------------GlyGlnGly 858
                                                                   |||
Db     142855 CAATTCCCTCCTGCTATGCAAGCTCCTGCTGCCGCACAAGCTCCTGTTGCTGCACAAGCT 142796

Qy        859 ThrThrAspGlyGluValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAla 878
              :::   |||   |||   ::::::|||:::::|||||||:::|||||||   |||
Db     142795 TCTGCTGAC---GAGGCAGAACGGAACAGGAAGTATCAGGCGACTCTTCAGTTGGCTCAG 142739

Qy        879 AsnLeuLeuLeuGlnIleGlnGlnLysGlnGlnGlnGlnSerSerGlyThrProAlaGly 898
                 ||||||   |||:::|||||||||          |||      |||   |||
Db     142738 CGCCTGTTGGGTCAGCTACAACAGAAGCCTGGAAATCAACCTTAAGGCTATCCTGGAAAT 142679

Qy        899 Gln 899
              |||
Db     142678 CAA 142676

RESULT 11
AP006756/c
LOCUS       AP006756              176553 bp    DNA     linear   PLN 28-JUL-2004
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 9,
            PAC clone:P0450E05.
ACCESSION   AP006756
VERSION     AP006756.2  GI:50725340
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Katayose,Y.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 9, PAC
            clone:P0450E05
  JOURNAL   Published Only in Database (2004)
REFERENCE   2  (bases 1 to 176553)
  AUTHORS   Sasaki,T., Matsumoto,T. and Katayose,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2004) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Jul 27, 2004 this sequence version replaced gi:46093787.
            Genes were predicted from the integrated results of the following:
            GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
            (http://www.softberry.com/), GeneMark.hmm
            (http://opal.biology.gatech.edu/GeneMark/), GlimmerM
            (http://www.tigr.org/tdb/glimmerm/glmr_form.html), RiceHMM
            (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
            (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4
            (http://globin.cse.psu.edu/html/docs/sim4.html), gap2
            (http://www.tigr.org/software/glimmerm/), BLASTN and BLASTX. The
            genomic sequence was searched against NCBI NonRedundant Protein
            database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
            sequence database at RGP or DDBJ. Protein homologies of the coding
            regions were searched against NCBI NonRedundant Protein database
            with BLASTP. ESTs represent the identified cDNA sequences using
            BLASTN with the corresponding DDBJ accession no. and RGP clone ID.
            Full-length cDNAs represent the identified cDNA sequences using
            BLASTN with the corresponding DDBJ accession no.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'.  A gene without
            significant homology to any protein but with full-length cDNA or
            EST homology (covering almost the entire length of partial
            sequence) is classified as an 'unknown' protein. A gene predicted
            by two or more gene prediction programs is classified as a
            'hypothetical' protein according to IRGSP standard. A gene
            predicted by a single gene prediction program is also classified as
            a probable 'hypothetical' protein and is included as a
            miscellaneous feature of the sequence.
            The orientation of the sequence is from T7 to SP6 of the PAC clone.
            This sequence of P0450E05 has an overlap with OSJNBa0065A15(DDBJ:
            AC137594) clone at 5' end and with OSJNBb0034B12(DDBJ: AP005735)
            clone at 3' end. Detailed information on overlap and assembly
            quality together with annotation of this entry is available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             Location/Qualifiers
     source          1. .176553
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="9"
                     /clone="P0450E05"
     gene            1479. .4246
                     /gene="P0450E05.1"
     mRNA            join(<1479. .1885,3329. .3366,3854. .3945,4100. .>4246)
                     /gene="P0450E05.1"
                     /note="start and end point are not identified"
     CDS             join(1479. .1885,3329. .3366,3854. .3945,4100. .4246)
                     /gene="P0450E05.1"
                     /note="similar to Arabidopsis thaliana chromosome 5,
                     At5g47870"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD34413.1"
                     /db_xref="GI:50725341"
                     /translation="MEALRLSGPAAPPFVPAARVAAAWGARRRGASRSSSSVVAKMEG
                     GGGGKGVPTTNYVVPLDKATGMTRPLVEILRDLNKRVPDKIIDPDTNTVPWSVLLPIP
```

```
                PSISSPARSALALQNDPADVAIGFRHSEVIELIGYHANRMLSFYAPGWCGEVRNVIYS
                DNGTVTVVYRVILRGTDGEAYREATGTAPVHEGRNDDAVAAAEEAAFCKACARFGFGL
                YLYHQDEIP"
gene            complement(5459. .5707)
                /gene="P0450E05.2"
mRNA            complement(<5459. .>5707)
                /gene="P0450E05.2"
                /note="start and end point are not identified"
CDS             complement(5459. .5707)
                /gene="P0450E05.2"
                /note="contains EST(s): C98104(C0709),D15486(C0709A)"
                /codon_start=1
                /product="unknown protein"
                /protein_id="BAD34414.1"
                /db_xref="GI:50725342"
                /translation="MSPTAVHQRFGVNFHRRWKPRTPRSSATVAKSKPTHLIYHVPLM
                LNHQIDGCCALDRCCPYEEEEAVVAMPYLMPPCGCSVD"
gene            6343. .8325
                /gene="P0450E05.3-2"
misc_feature    6343. .8325
                /gene="P0450E05.3-2"
                /note="contains full-length cDNA(s): AK066925
                non-coding transcript
                probably inactive due to including stop codon(s) in CDS"
gene            join(6475. .6690,7257. .7368,7494. .7565,7650. .7870,
                7958. .8359)
                /gene="P0450E05.3-1"
mRNA            join(6475. .6690,7257. .7368,7494. .7565,7650. .7870,
                7958. .8325)
                /gene="P0450E05.3-1"
                /note="supported by full-length cDNA(s): AK061814"
mRNA            join(6515. .6690,7257. .7368,7494. .7565,7650. .7870,
                7958. .8359)
                /gene="P0450E05.3-1"
                /note="supported by full-length cDNA(s): AK104595"
CDS             join(6581. .6690,7257. .7368,7494. .7565,7650. .7870,
                7958. .8063)
                /gene="P0450E05.3-1"
                /note="contains EST(s): D45991(S10331),AU091472(S10331)
                contains full-length cDNA(s): AK061814,AK104595
                similar to Arabidopsis thaliana chromosome 5, At5g11680"
                /codon_start=1
                /product="unknown protein"
                /protein_id="BAD34415.1"
                /db_xref="GI:50725343"
                /translation="MAENPQLFGNGMLVPFYGEMFVLARDGVEFHVDKIPSAPGGHAK
                TKGTIYLSNIRMVFVASKPVGNFFAFDMPLLYVHGEKFNQPIFHCNNISGFVEPVVPE
                NQNRALYSTHTFKILFKEGGCGTFVPLFLNLVASVRRYNQFEAQSAASMAPRVDPLQA
                VQTPVDDMMRHAYVDPNDPTKIFLQQPAPESQLRRRNYHGRADNAY"
gene            complement(join(8865. .8962,9022. .9136,9350. .9394,
                9516. .9536))
                /gene="P0450E05.4"
misc_feature    complement(join(8865. .8962,9022. .9136,9350. .9394,
                9516. .9536))
                /gene="P0450E05.4"
                /note="hypothetical ORF
                predicted by GeneMark.hmm
                this category is not included in IRGSP standard"
gene            11724. .14842
                /gene="P0450E05.5"
mRNA            join(11724. .11969,12522. .12640,12754. .12963,13048. .13196,
                13274. .13445,14009. .14101,14208. .14347,14438. .14726)
                /gene="P0450E05.5"
                /note="supported by full-length cDNA(s): AK062078"
mRNA            join(11729. .11969,12522. .12640,12754. .12963,13048. .13196,
                13274. .13445,14009. .14101,14208. .14347,14438. .14842)
                /gene="P0450E05.5"
                /note="supported by full-length cDNA(s): AK121370"
mRNA            join(11729. .11969,12522. .12640,12754. .12963,13048. .13196,
                13274. .13445,14009. .14101,14208. .14347,14438. .14788)
                /gene="P0450E05.5"
                /note="supported by full-length cDNA(s): AK098858"
CDS             join(11963. .11969,12522. .12640,12754. .12963,13048. .13196,
                13274. .13445,14009. .14101,14208. .14347,14438. .14537)
                /gene="P0450E05.5"
                /note="contains EST(s):
                D40583(S2655),AU183211(E50980),AU030382(E50980),AU091906
                (E10040),AU108323(S2655)
                contains full-length cDNA(s): AK062078,AK121370,AK098858"
                /codon_start=1
                /product="putative TGB12K interacting protein 3"
                /protein_id="BAD34416.1"
                /db_xref="GI:50725344"
                /translation="MEDQKKNAKPEGSSGSQRGAPPAPDAGLPNPFDFSQFSNLLNDP
                SIKEMAEQIASDPVFTQMAEQLQKSAHVTGEQGGPALDPQQYMETMTQVMQNPQFMSM
                AERLGNTLMQDPGMSSMLESLTSPSHKELLEERMSRIKEDPSLKGILDEIESGGPSAM
                VKYWNDPEVLQKIGQAMSINFPGDAATSTTLSGPEETEEDGGDDDESIVHHTASVGDA
                EGLKKALEDGADMDEEDAEGRRALHFACGYGELKCAEILLEAGAAVNALDKNKNTPLH
                YAAGYGRKECVDLLLKHGAAVTPQNLDGKTPIEVAKLNNQDEVLKVLEMDAFL"
gene            15883. .20771
                /gene="P0450E05.6-2"
misc_feature    15883. .20771
                /gene="P0450E05.6-2"
                /note="contains full-length cDNA(s): AK120151
                non-coding transcript
                probably inactive due to including stop codon(s) in CDS"
gene            join(17539. .17575,17680. .17780,18263. .18439,19056. .19213,
                19594. .19648,19740. .19793,20322. .20771)
                /gene="P0450E05.6-1"
mRNA            join(17539. .17575,17680. .17780,18263. .18439,19056. .19213,
                19594. .19648,19740. .19793,20322. .20771)
                /gene="P0450E05.6-1"
                /note="supported by full-length cDNA(s): AK060173"
CDS             join(17769. .17780,18263. .18439,19056. .19213,19594. .19648,
                19740. .19793,20322. .20492)
                /gene="P0450E05.6-1"
                /note="contains EST(s): C28886(C62721)
                contains full-length cDNA(s): AK060173"
                /codon_start=1
                /product="lecithin cholesterol acyltransferase-like"
                /protein_id="BAD34417.1"
                /db_xref="GI:50725345"
                /translation="MHQLLVECPSIYEMLPNPHFKWKKAPVVQVWRKKPEKDGIAELV
                LYEATDCLSLFQEALRNNELKYNGKTIALPFNMSVFKWATETRRILEDAELPDTVSFY
                NIYGTSYDTPYDVCYGSESSPIGDLSEVCHTMPVYTYVDGDGTVPIESTMADGFAAKE
                RVGIEADHRGLLCDENVFELLKKWLGVKEESTRRRRLSKSKVTDFAPS"
gene            complement(21894. .25877)
                /gene="P0450E05.7"

Alignment Scores:
Pred. No.:              2.38e-34        Length:         176553
Score:                  1117.50         Matches:        325
Percent Similarity:     45.76%          Conservative:   134
Best Local Similarity:  32.40%          Mismatches:     280
Query Match:            23.28%          Indels:         264
DB:                     8               Gaps:           39

US-09-920-705-3 (1-901) x AP006756 (1-176553)

Qy         91 AlaLysProCysLysSerLeuTrpValGlyGlyIleGlyProAsnValSerLysAspAsp 110
              |||:::     |||:::||||||||||||||||||      ::::::||||||:::::: 
Db     163887 GCTAGAGCTGTTAAGAACTTATGGGTTGGTGGCATTAGCTCCTCAATTTCAAAGGAGGAG 163828

Qy        111 LeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArg 130
              ||||||||||||||   ||||||||||:::::      |||   |||::::::
Db     163827 CTGGAGGAGGAGTTTAAGAAATTTGGAAAGGTTGATGGTATTGCATTCTCTCGTGATCAG 163768

Qy        131 LysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGlnAla---LysSerMet 149
              :::|||:::||||||:::   :::::::::||||||:::   |||   :::::::::
Db     163767 ACTTCCGCGTACATTGATTTTGACAAGCTGGAAGATGCAATTTCGGCACATAGAGCCTTG 163708

Qy        150 AsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeuArgSerGln-AlaPr 169
              ||||||:::   :::||||||      |||   |||||||||   ||||||:: |||
Db     163707 AATGGAAGAGTTTTAGGTGGACAAGAACTGTGTGTTGATTTCCAAAGGTCCAGGGGCAGA 163648

Qy        169 oLysLysGlu---------------------------------------------- 172
```


Db 163647 GCGGTGTGTATCCATTAATTATATGTATTGATATACTGTTTGCATTTTCTTGGTAACAAG 163588
Qy 173 ----------GlnTrp------Al 175
Db 163587 CTTATGAATGTTCTTGTTGCATGTAATATTTCTGAATTTTTTTTAGGAATGGTTGGAGAC 163528
Qy 175 aGlySerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGlnTyrProHis----- 193
Db 163527 AGGCAGCTTCAATGGTAGAACTGGA----------CCTGCAAAGGT 163492
Qy 193 ---------- 193
Db 163491 ATTTTCTGTTTGCATTTCCTTTCTTGTATTACTGGTTGTTTTTCTGTTATATCTGCTGAT 163432
Qy 194 ----------SerTyrGlu----------AspPheLysGlyAs 201
Db 163431 GTTGCTAACTATCGTTTCATATTCTGTTGGCATGTCAGGGTTATGGAGTACGAAATCGAG 163372
Qy 201 pValGln-ProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnA 221
Db 163371 AATCCAATCCCACCAATGTTCTTTGGGTAGGTTTCCCTAATACAGCTAAA------ATCA 163318
Qy 221 spGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerT 241
Db 163317 ATGAGGAGGCACTTCGGCAAGCGATGGCTGTACATGGTGCTGTTACAAATACAAAAGTTT 163258
Qy 241 yrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysL 261
Db 163257 TCCCAACAAGGCAATATGCTTTTGTGGAGTTTGCAACAGTTGGGGAAGCTTCTAATGCTA 163198
Qy 261 ysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsnA 281
Db 163197 AGAAAAATCTAGATGGTCGTCTTTTCAATGATCAGAGAATTCAAATTCTTTTCTCGAACA 163138
Qy 281 spGluLeuProProGluGln---AspAspThrSerPheTyrSerGlyMetLysArgSerA 300
Db 163137 GTGAGCTTGCGCCAAATAAATTAGATAATCCAACAGCAGTTTCTGGATTTCCAAAATCA- 163079
Qy 300 rgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIleP 320
Db 163078 -----GAGATGTATTATGATGATGGTCAGTATGGCGCTTCTGATTAT---------- 163037
Qy 320 roGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrA 340
Db 163036 --------TTTGATCCTCGACGGGGAAGATCAAGATAC----------TTTGAGTACA 162997
Qy 340 snAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleL 360
Db 162996 GTGGTGTC---------- 162989
Qy 360 euProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgProM 380
Db 162988 --------CCTGTTTCTGGTGGTATTCTTCCATCACCGGAATCTGGA------AATCCTC 162943
Qy 380 etArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuValArgGluSerL 400
Db 162942 TCTTAACTGGACGATCTGCTCAGAGTACATTCGATCCA----------AGAGAAGCCA 162895
Qy 400 ysArgThrArgArgAspGlySerValAsp----------GlyP 411
Db 162894 AGAGGTTGCGGTTGGATGCTGCAGCTGACCCTTATGATACAAGGGCAGGTAGTGAGGGTC 162835
Qy 411 heThrProMetGlyValAspGluArgSerPheGlyArgGlySerValAlaAlaArgPro- 430
Db 162834 TTTATTCTGCTGGGTACTCTCAGCGCGAGAGTGCACGTTCTGAGAGGAGCTCAAGTCCTG 162775
Qy 431 --------IleArgGlyProProAspSer----------AspHisIleTrpArgGlyM 444
Db 162774 CTATTCGAATCCATGGCACAGTGCACCGAACATCATACCTTGAGCACTTCTGGCGTGGTA 162715
Qy 444 etIleAlaLysGlyGlyThrProValCysCysAlaArgCysValProMetGlyLysGlyI 464
Db 162714 GCATTGCCAAAGGTGGATCTCCTGTATGTCGTGCTCGCTGTTTGCCTATAAGGAAGGGTG 162655
Qy 464 leGluThrLysLeu---------- 468
Db 162654 TTGAGATACCTTTGTATGTATACACTCACCATCTTAATCTTGTTCTCTACAGTTGATTTC 162595
Qy 469 ----------ProGluValValAsnCysS 475
Db 162594 TATAGCCTCTGGATCTTACACATATTCCACTGTTTTGCAGACCGGATGTTGTTAATTGTT 162535
Qy 475 erAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGluI 495
Db 162534 CCGCTAGAACTGGACTAGATATGCTGGCAAAGCATTATAGAGACGCTTCAGGGTTTGATA 162475
Qy 495 leValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgT 515
Db 162474 TCGTCTTCTTCTTGCCAGATAGTGAAGATGACTTTGTTTCTTACACTGAATTTTTGCGCT 162415
Qy 515 yrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheL 535
Db 162414 ACTTAGGCTCAAAAAGTCGGGCAGGGGTTGTAAAGGTTGATGGGGGGACTACTTTATTTT 162355
Qy 535 euValProProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuT 555
Db 162354 TGGTCCCACCTTCCGATTTTTTACGAAATGTTTTGCAAGTTGATGGTCCAGAACGTCTTT 162295
Qy 555 yrGlyValValLeuLysLeuPro----------ProProAlaValProValT 569
Db 162294 ACGGTGTAGTATTACACATTCCGCAAATGTCTGCTGCTGCTCCTGCTTCTGCTCCCACTC 162235
Qy 569 hrAlaSerTyrArgGlnGlu---------SerGlnSerAsnProLeuHisTyrMetAspG 586
Db 162234 CTGCTGTACAAAGGCCACAACTAACTGCACCAGAATCACAACCT------TACTACGATG 162181
Qy 586 lnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGluAsnTyrI 606
Db 162180 AA---AGGGAAATTCCTTTGCAGAGAAGGTATAGTATGATTACGCCTAGCAACAATCACC 162124
Qy 606 leArgGlyAlaProGluHisLeuThrAla---------- 615
Db 162123 ATCGAGATGCT---GATCACCGTGGATCTTTACGTGAGGATTCATTGCATCAATTGGGGC 162067
Qy 616 -----AlaSerLysProSerValSerGluPro----------LeuArgI 627
Db 162066 AAATTCTTGCACGACCTCGGGTGGATGAAGGACAAGTAGTTCAGCCAAACCTTGCCGGTA 162007
Qy 627 leProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAlaThrL 647
Db 162006 TCCCTACAAACGCAGGACTACAGGTACAGCCTTCACTCCAGCCTGATATGATAGCCACTT 161947
Qy 647 euAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMetS 667
Db 161946 TAGCTAAACTTTTGCCAAGTGGGCAGTCATCAGCTCTGGTCACTGGCCAGCTTCCCCTGA 161887
Qy 667 erGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyGluA 687
Db 161886 GC------TCAACAGACCAACCTGCATTAACCCAAATGAAT---------GATGCATCTA 161842
Qy 687 laProSerGlnAlaTrpLysArgGlyProGlnThrValHisAspAlaSerAsnGlnSerP 707
Db 161841 CTCTTGCAAAAGTATGGAGA------CCTGAAAATCAAGCAATGGCTTCTACTTCATCCT 161788
Qy 707 heGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeuProProProProSerArgT 727
Db 161787 TAGAGCAGATCGGTAAT---TTCCAGCATTCAGGACAA---------CAGTTCAGCAAAC 161740
Qy 727 yrProProAlaSerAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMetGlnT 747
Db 161739 AAGCTGGAGCAGTTCATCTGCCAAACTAT------GGGAACTTGGCAGGTGCACAGGAGC 161686
Qy 747 yrGlnSerGln----------SerValAsnMetProGlnLeuSerP 759
Db 161685 ACCCAACACAACATAGTGCTTACAATCCTGAAATGACTCTAAACTTGCCACCACCACCAC 161626

```
Qy        759 roLeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisP 779
              ||   |||   :::|||   ::::::     :::   :::|||        :::    |
Db     161625 CACCTCCTACGCTGCCACCCAGTTCTGCTATATTATCATCGCAAGTTGGACACAGTTTGC 161566

Qy        779 roValSerGlnProMetValGlnGlnTyrGlnProGluAla---SerMetProAsnGlnA 798
              ||   :::|||          ||||||||||||||||||          |||         |
Db     161565 CT---ACGCAGATGAGTCAGCAACAGTATCAACCAGAGCAGTATTATATGACTCAAAGCA 161509

Qy        798 snTyrGlyProIleProSerTyrGlnGlnAlaAsn-PheHisGlyVal------------ 813
              ||||||||   :::   :::          :::||| |||
Db     161508 ATTATGGTCAATTGGCCACAGTTAGCAGTTCTAATCTTCAGGTCAGCAATTCTAACATCC 161449

Qy        814 --------------------------------ThrThrAsnGlnAlaGlnAsnLeuAsn 822
                                             |||   ||||||       |||
Db     161448 CATTTCTTCCTTATTCCCAAGTAAACCTTGCCAACAAATAATCAGATAGGGTTTTTG--- 161392

Qy        823 ProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsn------------ 838
                 :::   ||||||   |||:::      |||   |||:::   |||
Db     161391 ---GCAGTGTTTCAGCATGCTTTGACTTTGCCTAATGATAGGGTAAACTATGGTTTCTCT 161335

Qy        839 LeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAsp--------- 855
              |||:::      :::|||   :::          |||
Db     161334 TTACAGGCGCACCATCAACAAATTGTTGCGACTCCTGCTGCACAAGCTCCTGTTGCTGCG 161275

Qy        856 ----------------------------------------------------GlyGlnGly 858
                                                                   |||
Db     161274 CAATTCCCTCCTGCTATGCAAGCTCCTGCTGCCGCACAAGCTCCTGTTGCTGCACAAGCT 161215

Qy        859 ThrThrAspGlyGluValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAla 878
              :::   |||   |||   ::::::|||::::::||||||:::|||||||||   |||
Db     161214 TCTGCTGAC---GAGGCAGAACGGAACAGGAAGTATCAGGCGACTCTTCAGTTGGCTCAG 161158

Qy        879 AsnLeuLeuLeuGlnIleGlnGlnLysGlnGlnGlnGlnSerSerGlyThrProAlaGly 898
              ||||||   |||:::|||||||||          |||     |||   |||
Db     161157 CGCCTGTTGGGTCAGCTACAACAGAAGCCTGGAAATCAACCTTAAGGCTATCCTGGAAAT 161098

Qy        899 Gln 899
              |||
Db     161097 CAA 161095
```

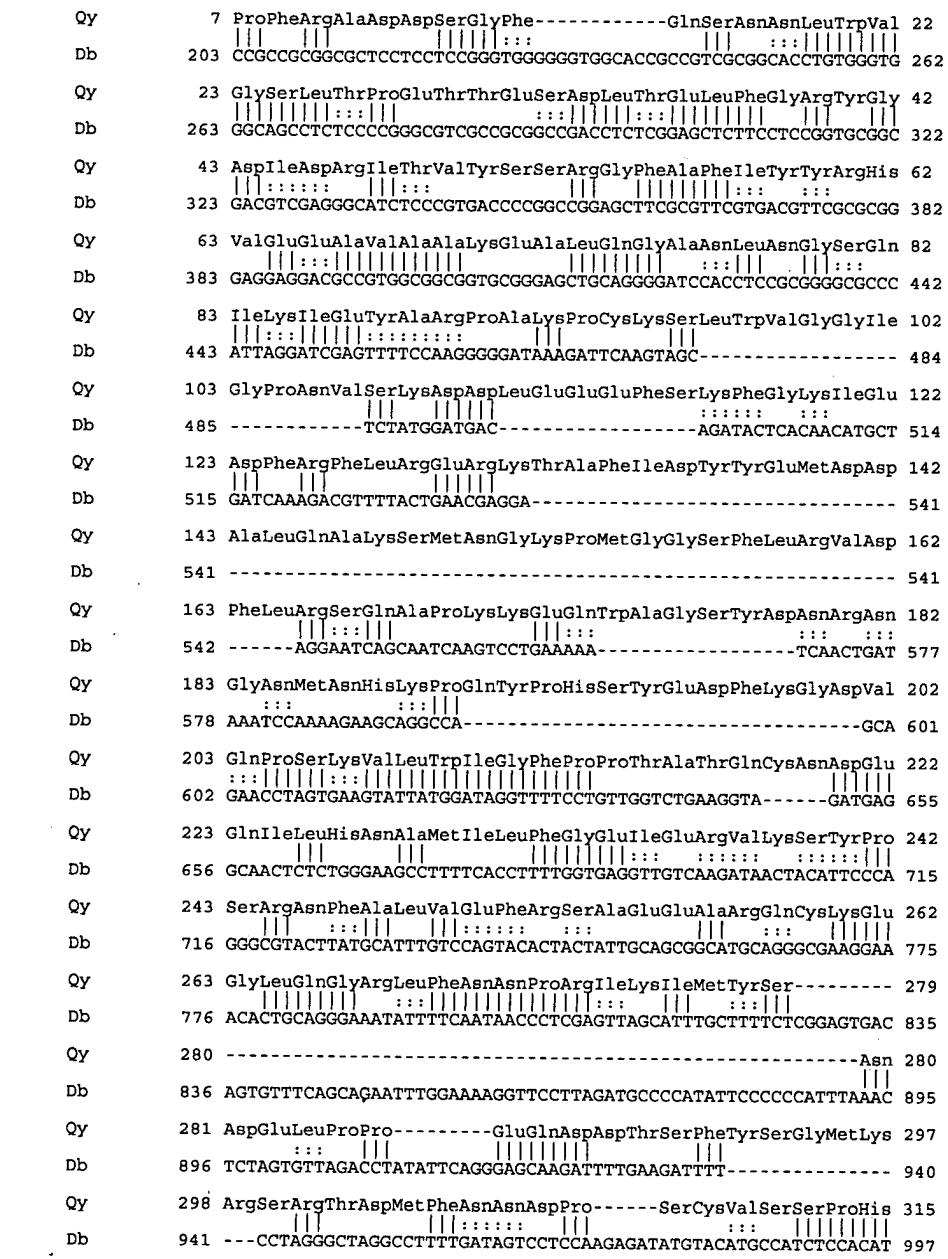
RESULT 12
AK100889
LOCUS AK100889 3455 bp mRNA linear PLN 24-JUL-2003
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J023129A05, full insert sequence.
ACCESSION AK100889
VERSION AK100889.1 GI:32986098
KEYWORDS FLI_CDNA; CAP trapper.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1
AUTHORS The Rice Full-Length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-Length cDNA Project Team:, Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K., Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y., Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M., Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J., Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J., Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:, Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y., Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Yoshino,M. and Hayashizaki,Y.
TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice
JOURNAL Science 301 (5631), 376-379 (2003)
MEDLINE 22752273
PUBMED 12869764
REFERENCE 2 (bases 1 to 3455)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M., Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT This clone is one of the 28K full-length cDNA clones from japonica rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.
FEATURES Location/Qualifiers
source 1. .3455
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J023129A05"
ORIGIN

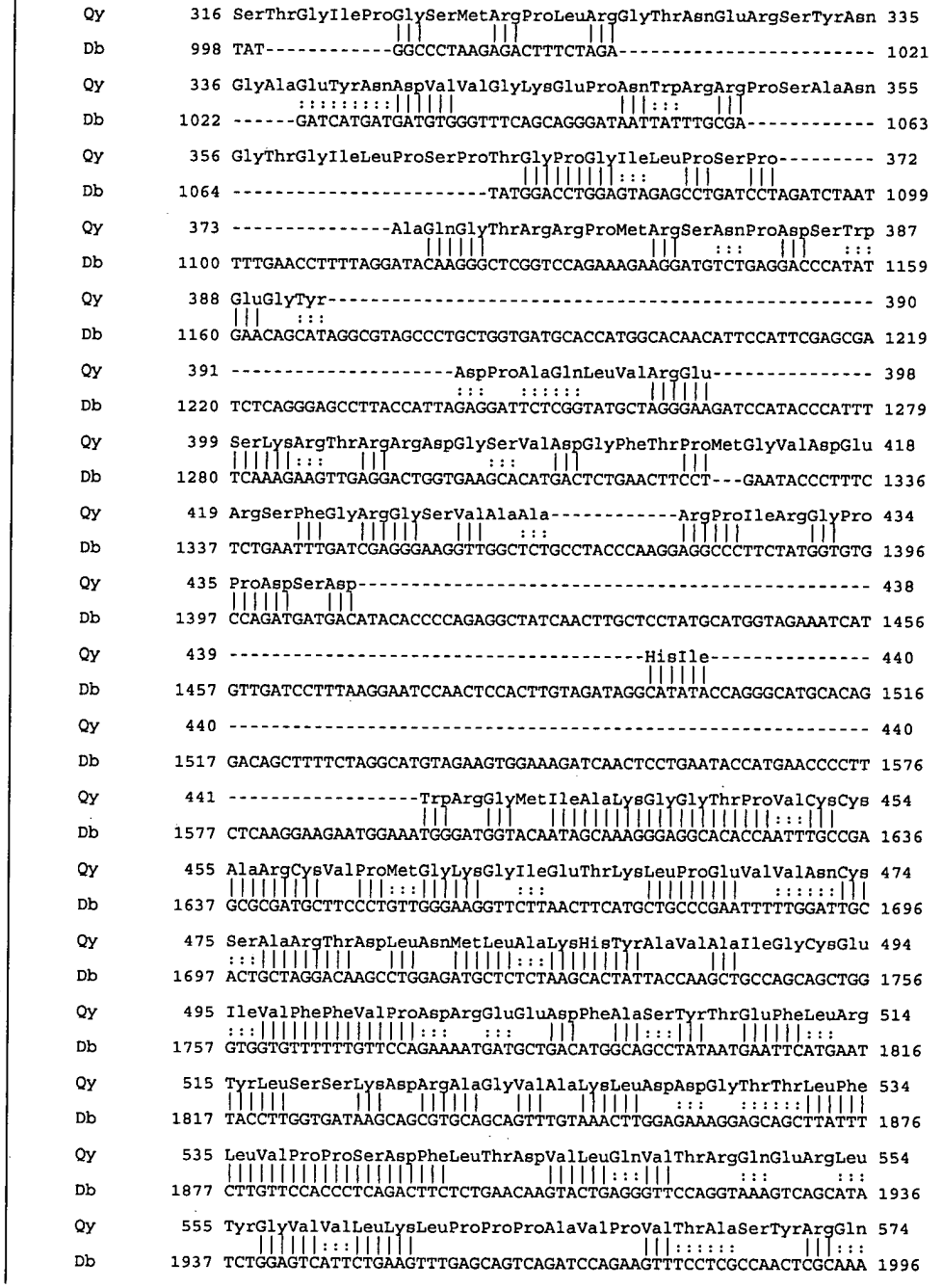
Alignment Scores:

Pred. No.:	2.05e-23	Length:	3455
Score:	799.00	Matches:	298
Percent Similarity:	37.20%	Conservative:	119

Best Local Similarity: 26.58% Mismatches: 326
Query Match: 16.64% Indels: 378
DB: 8 Gaps: 45

US-09-920-705-3 (1-901) x AK100889 (1-3455)





```
Qy       575 GluSer---------------------------------------------------- 576

Db      1997 CCAGAAACATTTGTGAGTCATTTGAACCATGATGTTCGTGCTCATGAGGATCTAGATGCA 2056

Qy       577 ------GlnSerAsnPro------LeuHisTyrMetAspGlnAlaArgAsp--------- 589
                 :::   ||||||      :::     :::   |||     |||
Db      2057 TTGAGAAGAATCAACCCACCAGATATCAGGCCACTTCCTCAGGGTTCAGATTATCTCGGG 2116

Qy       590 ---SerProAlaAsn---------AlaSerHisSerLeuTyrProProArgGluAsnTyr 605
                |||||   ::         ||||||      |||   ||||||   :::
Db      2117 TTGTCGCCTGGAAGCTATAATCCAGCAAGTGCACATTTGGTTCCGCCTTACAAG------ 2170

Qy       606 IleArgGlyAlaProGluHisLeuThrAla----------------------------- 615
                    ||||||   :::|||   :::
Db      2171 TTTGGAAATGCTCCTTCATATCTAGAATCTGAATTAGCTCATCAAAAGCATCCACCTGAC 2230

Qy       616 --------------AlaSerLysProSerValSerGluProLeuArgIleProAsn--- 629
                           |||                 |||      :::|||:::
Db      2231 TCCCACAGGGAGATAGCACATGACAAGCAGCAGCAACACCCAGATGTATTGCCCTCAAGA 2290

Qy       630 -----------------AsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeu 643
                          |||   :::|||   :::|||                     |||
Db      2291 TGGTCAGATAACATTTACAATCCAAGTCCAGGTTCTGGA--------------AATTTG 2335

Qy       644 LeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHis 663
                       |||   :::|||   ||||||                |||||::::::
Db      2336 AATTATTTGGCTGAGAGTGCGATCCCACATACATCAACTGATAGGACACCAGAGGCATAC 2395

Qy       664 Gln------ProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsnGly 681
                 |||         |||         |||||||||
Db      2396 TCATTTGCTCCTCAAGGAGTACCAAAA------GTGAGTACA------------------ 2431

Qy       682 LeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHisAsp 701
                :::|||   ||||||                                      :::
Db      2432 ------TCAGGGTATGCACCA-------------------------GTTGCAGATGAG 2458

Qy       702 AlaSerAsnGlnSerPhe------GlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGln 719
                 |||||||||   |||:::         |||     :::|||              |||
Db      2459 GCATCAAACATGTCCTACCCTCCCATGCAACCTGCATCACAGCAGGTAGTTAGACCTCAA 2518

Qy       720 LeuProProPro---ProSerArgTyrProPro-------------------------- 729
                 ||||||    |||         ||||||
Db      2519 CAACCTCCATCTCTCCCATTATCGCTTCCACCAGAGCAACTTGCACAATTGGCCACTCTT 2578

Qy       730 ------AlaSerAsnAsnPro-----------------------AsnTyrThrSerGly 739
                    |||   |||                          |||      ||||||
Db      2579 CTTGCACAACAAAACCAACCAGGAAAAGAGCCTGTGGACAGCTTAAACAAAGAATCAGGA 2638

Qy       740 MetVal---------HisGlyAsn---------MetGlnTyrGlnSerGlnSerValAsn 753
                 :::         ||||||:::         |||   :::   |||   |||:::
Db      2639 TTCATACGGAATCCACATGGACATTCTTCAATGATGCCACACAGCTCAGGTTCTATC--- 2695

Qy       754 MetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGln 773
                 |||      :::   ||||||      |||
Db      2696 ---CCTGTCCAAAACTCATTGCCACCTGCTCCG----------------------CCA 2728

Qy       774 GlySerSerAsnHisProValSerGlnProMetValGln-------------------- 786
                 :::|||         |||      |||   ||||||
Db      2729 TCTGCATCACAGTTACAGGTTCATGCGCCACCAGTTCAAGGTTCAGTGCCACCAAATCCA 2788

Qy       787 ------GlnTyrGlnProGluAlaSerMetProAsnGlnAsnTyrGlyProIleProSer 804
                    :::   |||   |||   ||||||:::   |||      |||:::|||
Db      2789 TCCATTATGCATACACCGAATGCTCCTATGCCTTCTCACAACACTTTGCCTTTACCCCCT 2848

Qy       805 TyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSer 824
                    :::                 |||   |||   :::      |||
Db      2849 ATGCATCCTTCGGGA----------------AATCCAGCTCATTCTTCCATGCCTTTG 2890

Qy       825 GlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGluProGlnAsnGln 844
                 :::                 |||                 :::      :::
Db      2891 AGATCGTTCGTCCCTCCTCTTCCTGAAGGCCCTCCACCCCTTAGGCAGCACACATCAAGT 2950

Qy       845 AlaLeuArgLeuGlnProMetIleSerGlyAspGlyGln----------------Gly 858
```

```
                |||||:::   ||||||   :::            |||
Db      2951 GCTCTACAAGCGCAACCTGCACTTCCCTCTGGACCACAGACTAGCCAGCAACCGTCTGCT 3010

Qy       859 ThrThrAspGlyGluValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAla 878
                    |||         |||      :::|||  |||:::||||||||   ||||||
Db      3011 CAGGAAGACCATCACGGAGACCCTCAAAAGCGCCTTCAAGCAACATTGCAATTGGCAGCA 3070

Qy       879 AsnLeuLeuLeuGlnIleGlnGlnLysGlnGlnGlnSerSerGlyThrProAlaGly 898
                 ||||||   ||||||||||||:::   :::                |||   |||
Db      3071 ACCCTACTTCAGCAGATACAGCAACAATCTAAA-----------------CCTGGTGGC 3112

Qy       899 Gln 899
                 |||
Db      3113 CAG 3115
```

```
RESULT 13
AX420424
LOCUS       AX420424               3715 bp    DNA     linear   PAT 18-JUN-2002
DEFINITION  Sequence 6 from Patent WO0212518.
ACCESSION   AX420424
VERSION     AX420424.1  GI:21524580
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Amasino,R.M., Schomburg,F.M., Michaels,S.D. and Patton,D.
  TITLE     Floral induction gene
  JOURNAL   Patent: WO 0212518-A 6 14-FEB-2002;
            WISCONSIN ALUMNI RESEARCH FOUNDATION (US)
FEATURES             Location/Qualifiers
     source          1. .3715
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
                     /note="FPA promoter plus intron"
ORIGIN

Alignment Scores:
Pred. No.:              1.78e-17        Length:         3715
Score:                  651.50          Matches:        171
Percent Similarity:     27.23%          Conservative:   0
Best Local Similarity:  27.23%          Mismatches:     0
Query Match:            13.57%          Indels:         457
DB:                     6               Gaps:           1

US-09-920-705-3 (1-901) x AX420424 (1-3715)

Qy         1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1832 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 1891

Qy        21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1892 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 1951

Qy        41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1952 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 2011

Qy        61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2012 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 2071

Qy        81 SerGlnIleLysIleGluTyrAlaArgPro---------------------------- 90
             ||||||||||||||||||||||||||||
Db      2072 AGTCAAATTAAGATCGAATACGCACGACCGGTTTGTTCTTATCTATATCTTCGTTTGTTC 2131

Qy        90 ------------------------------------------------------------ 90

Db      2132 TCTAACTTTGATTGTCTTTTGTCAACGATTATACTCTTTTTGCGAATTCATAGTCCAGGT 2191
```

```
Qy        90 ------------------------------------------------------------ 90
Db      2192 TCACAAAACTTTGATGATGCTTGTTTAGTCCAAAAATTTCTTGTTGAATCTGTTTTTTTT 2251
Qy        90 ------------------------------------------------------------ 90
Db      2252 TTCCTCATCGTACAAATCAAAGTCGAAACCTAGTTTTTTTCTATTATACGTCGTTAGCTT 2311
Qy        90 ------------------------------------------------------------ 90
Db      2312 AAGGCGAAACCTGATCCGATCGAAACGTCTTTTCTCAAATTACTTTGGTTATATCGAACT 2371
Qy        90 ------------------------------------------------------------ 90
Db      2372 CGCGCAAAGCCAAACCACAGAGAAGCTCTGCAAAATTTGATGTTAAAGCATATATAACTC 2431
Qy        90 ------------------------------------------------------------ 90
Db      2432 TTAGCGAATGAGCTCTGCAAAAGAATACATCAAACACATGTTTACTCTCGTTTATGCGAA 2491
Qy        90 ------------------------------------------------------------ 90
Db      2492 GAAGGTTTAATCCGATTGTCGTTTATCTGCAAACTTGTATCCGCGTTTAGCAGTCTGATT 2551
Qy        90 ------------------------------------------------------------ 90
Db      2552 TCAGGTTCGCTTCAGATGTTAAATCTCACAAGCTTGAGTATGAATGTATTGCGTACTCCA 2611
Qy        90 ------------------------------------------------------------ 90
Db      2612 CTCGAAATTCGCAATGGTGAAAATTTGGAAGCGAGCAAACTCTATCATCGGCCAAACAGA 2671
Qy        90 ------------------------------------------------------------ 90
Db      2672 TAAGAAATTTTGGAGTTTAAAAGTTCAGTTTCCTGCAAAAATCAAACCGCGTGGAGAAAT 2731
Qy        90 ------------------------------------------------------------ 90
Db      2732 TTGTCTATGGCGGTACGTAGATATCAATAACTGCATCTGCGAAACAGAAAATTCTGGCAA 2791
Qy        90 ------------------------------------------------------------ 90
Db      2792 CCGTATCGTCCTAACTCCTAGTATCGCTGGCACATATCCATATCAGTGTGAGGGAGTTTT 2851
Qy        90 ------------------------------------------------------------ 90
Db      2852 GGGCTTGTTTTGGCTCGATGCTTCAGGAAGACAAATTACGTGGTTGTTAAGGCGGCTAAC 2911
Qy        90 ------------------------------------------------------------ 90
Db      2912 TCTACCAATCAGAAACGCTTTATTCGAAGAACCATGTTTGTTCCTCAATTCCCATCCCTA 2971
Qy        90 ------------------------------------------------------------ 90
Db      2972 CGTACAATCTGGGCTTTCCCATTGTAGTTCCTTAGGAAGTTGACTTCTTCACAACCATTC 3031
Qy        90 ------------------------------------------------------------ 90
Db      3032 TGGGATTGGTACAGTTGCAGTGGACAAGACAAATATTCATTTGCAGGCAGACTCAAATTT 3091
Qy        90 ------------------------------------------------------------ 90
Db      3092 CAATGTCTTGCCGGCTTTGAGTACTTAAATGGAGTGTTCAGGGATTGGTTTATTTGGGAC 3151
Qy        90 ------------------------------------------------------------ 90
Db      3152 TCAGGCAATGGATAAAGACAGGAATGTTTTGCAGAAAGTATTAATGTCTTTCCGGCTTTG 3211
Qy        90 ------------------------------------------------------------ 90
Db      3212 GGGACTACATACTTCTCCTACAGACAAAATCTGATGTTTAGGCAAGAGAAACTATTAACA 3271
```

```
Qy        90 ------------------------------------------------------------ 90
Db      3272 CTGATTTAATAGAGAAAGAGGAGATGGTTTCTCCTTGCGGCAATTTTATTTGTTTAGGAA 3331
Qy        90 ------------------------------------------------------------ 90
Db      3332 AGCAATTGATATGAATTGGTGTCGTAGTGTAGTTGAAATTACTAGTTAGTTTGTGTGTTT 3391
Qy        90 ------------------------------------------------------------ 90
Db      3392 AGTTTCCTTGATGTTTGATGCTTTATTCTTGGCAACCTATCTGGGTAGTATCGCCTTCTT 3451
Qy        91 ----------------------AlaLysProCysLysSerLeuTrpValGlyGlyIleGly 103
                                    ||||||||||||||||||||||||||||||||||||||
Db      3452 ATCGACCTTTTCTTGTTGCAGGCAAAACCTTGTAAGAGTCTATGGGTGGGTGGAATCGGC 3511
Qy       104 ProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAsp 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3512 CCTAATGTCTCCAAGGATGACCTGGAGGAAGAGTTCAGCAAGTTTGGGAAAATCGAGGAT 3571
Qy       124 PheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAla 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3572 TTTAGGTTTCTCAGAGAACGCAAGACAGCTTTCATTGATTATTATGAGATGGATGATGCT 3631
Qy       144 LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPhe 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3632 TTACAGGCTAAGAGCATGAATGGAAAGCCTATGGGTGGTAGCTTTTTGCGTGTTGATTTT 3691
Qy       164 LeuArgSerGlnAlaProLysLys 171
              ||||||||||||||||||||||
Db      3692 CTCCGGTCACAAGCGCCAAAAAAA 3715
```

```
RESULT 14
AP004011/c
LOCUS       AP004011              130961 bp    DNA     linear   PLN 28-JUL-2004
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 9,
            BAC clone:OJ1381_H04.
ACCESSION   AP004011
VERSION     AP004011.3  GI:50725505
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 9, BAC
            clone:OJ1381_H04
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 130961)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Jul 27, 2004 this sequence version replaced gi:46092388.
            Genes were predicted from the integrated results of the following:
            GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
            (http://www.softberry.com/), GeneMark.hmm
            (http://opal.biology.gatech.edu/GeneMark/), GlimmerM
            (http://www.tigr.org/tdb/glimmerm/glmr_form.html), RiceHMM
            (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
            (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4
            (http://globin.cse.psu.edu/html/docs/sim4.html), gap2
            (http://www.tigr.org/software/glimmerm/), BLASTN and BLASTX. The
            genomic sequence was searched against NCBI NonRedundant Protein
            database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
            sequence database at RGP or DDBJ. Protein homologies of the coding
            regions were searched against NCBI NonRedundant Protein database
            with BLASTP. ESTs represent the identified cDNA sequences using
```

BLASTN with the corresponding DDBJ accession no. and RGP clone ID.
Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no.
A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.
The orientation of the sequence is from M13rev to -21M13 of the BAC clone. This sequence of OJ1381_H04 clone has an overlap with P0592C05 (DDBJ: AP004756) clone at 5' end and an overlap with OSJNBa0062A09 (DDBJ: AP004736) clone at 3' end. The sequence was generated by combining Monsanto and RGP-Japan sequencing data.
Detailed information on overlap and assembly quality together with annotation of this entry is available at http://rgp.dna.affrc.go.jp/GenomeSeq.html.

FEATURES Location/Qualifiers

source 1. .130961
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="9"
/clone="OJ1381_H04"

gene complement(3074. .6118)
/gene="OJ1381_H04.1"

mRNA complement(join(3074. .3542,3637. .3762,4173. .4312, 5966. .6063))
/gene="OJ1381_H04.1"
/note="supported by full-length cDNA(s): AK104965"

mRNA complement(join(3105. .3542,3637. .3762,4173. .4312, 5966. .6118))
/gene="OJ1381_H04.1"
/note="supported by full-length cDNA(s): AK073729"

CDS complement(join(3402. .3542,3637. .3762,4173. .4312, 5966. .6005))
/gene="OJ1381_H04.1"
/note="contains EST(s): AU174422(S6559),C25488(S6559) contains full-length cDNA(s): AK104965,AK073729"
/codon_start=1
/product="putative ubiquitin-conjugating enzyme family protein"
/protein_id="BAD32975.1"
/db_xref="GI:50725506"
/translation="MTLGSSGAGSSVVVPRNFRLLEELERGEKGIGDGTVSYGMDDAD DIYMRSWTGTIIGPHNSVHEGRIYQLKLFCDKDYPEKPPSVRFHSRINMSCVNHETGA VDSRKFGVLANWQREYTMETILTQLKKEMATPQNRKLVQPPEGTFF"

gene complement(8645. .9266)
/gene="OJ1381_H04.2"

misc_feature complement(8645. .9266)
/gene="OJ1381_H04.2"
/note="probably inactive due to 3'exon missing in CDS pseudogene, serine/threonine phosphatase PP7 -related"
/pseudo

gene complement(9412. .10109)
/gene="OJ1381_H04.3"

misc_feature complement(9412. .10109)
/gene="OJ1381_H04.3"
/note="probably inactive due to 3'exon missing in CDS probably inactive due to 5'exon missing in CDS pseudogene, transposase related protein"
/pseudo

gene complement(14903. .15292)
/gene="OJ1381_H04.4"

mRNA complement(<14903. .>15292)
/gene="OJ1381_H04.4"
/note="start and end point are not identified"

CDS complement(14903. .15292)
/gene="OJ1381_H04.4"
/note="similar to Oryza sativa chromosome 7, P0669H03.22"
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD32976.1"
/db_xref="GI:50725507"
/translation="MASRMVAATAPKTSHRGCDRGQRWRWRLPSPPTRMASRIAAAAC SEPSRVDGVEDGCSRASCYRWQRRRRLPSLLPWIGPMISATAVAPEAFTEVVDDGGSD SGGRRQWWRERKRKEKRWEGGEDETDT"

gene join(16127. .16171,16261. .16372,16416. .16460,16533. .16753)
/gene="OJ1381_H04.5"

misc_feature join(16127. .16171,16261. .16372,16416. .16460,16533. .16753)
/gene="OJ1381_H04.5"
/note="hypothetical ORF predicted by GENSCAN this category is not included in IRGSP standard"

gene 16879. .18432
/gene="OJ1381_H04.6"

mRNA <16879. .>18432
/gene="OJ1381_H04.6"
/note="start and end point are not identified"

CDS 16879. .18432
/gene="OJ1381_H04.6"
/codon_start=1
/product="putative monosaccharide transporter 6"
/protein_id="BAD32977.1"
/db_xref="GI:50725508"
/translation="MVAASGAGRPEYPGGLTMFVSMACLVAATGGLIFGYDIGVSGGV TSMDPFLSRFFPSVYRAQSAAAAAAGGNQYCRFDSQLLTMFTSSLYLAALASSLGAAT VTRVAGRKWSMFAGGLVFLAGCALNGAAANVAMLIVGRVLLGVGIGFANQSVPVYLSE MAPARMRGMLNNGFQMMITTGVLAANLINYGTARIAGGWGWRLSLALAAVPAAVMTAG ALFLPETPNSLLERGRRGEARRMLQRVRGEGVDMEDEYNDLVAAGEASHAVASPWRDI LRRRNRPPLVMAVAIPLFQQLTGINVIMFYAPVLFRTLGFGGGASLMSAVITGGVNMA ATLVSVLAVDRVGRRALFLEGGAQMVASQAAVGALIGARLGWSGTAAIPAGYAAAVVA AMCVYVAAFAWSWGPLAWLVPSEVMPLEVRPAGQSITVAVNMAMTFAVAQAFLPLLCR LRFVLFFFFAGWVAAMTAFVALFVPETKGVPIEDMAAVWSDHWYWKRFVDGDGDGARR RGDIEMGHK"

gene complement(20105. .23575)
/gene="OJ1381_H04.7"

mRNA complement(join(20105. .20513,20597. .20726,20898. .20966, 21052. .21171,21461. .21545,21670. .21755,21867. .22019, 22113. .22235,23079. .23575))
/gene="OJ1381_H04.7"
/note="supported by full-length cDNA(s): AK071036"

CDS complement(join(20437. .20513,20597. .20726,20898. .20966, 21052. .21171,21461. .21545,21670. .21755,21867. .22019, 22113. .22235,23079. .23462))
/gene="OJ1381_H04.7"
/note="contains EST(s): AU100795(C51066),C27141(C51066) contains full-length cDNA(s): AK071036"
/codon_start=1
/product="phosphoenolpyruvate/phosphate translocator"
/protein_id="BAD32978.1"
/db_xref="GI:50725509"
/translation="MQSAAAVGLLRPCGATTAAAPLQLRNPSPRGFGVGVGQPLLPPR GLRLSAVAPRAGISARRIGLVPASPEQEDERRRGARDVAVAATAAAAGEAGAEEGGGL AKTLQLGALFGLWYLFNIYFNIYNKQVLKVFPYPINITNVQFAVGTVIALFMWITGIL KRPKISGAQLAAILPLAMVHTMGNLFTNMSLGKVAVSFTHTIKAMEPFFSVLLSALFL GEMPTPFVVLSLVPIVGGVALASLTEASFNWAGFWSAMASNVTFQSRNVLSKKLMVKK EESLDNITLFSIITVMSFFLLAPVTLLTEGVKVTPTVLQSAGLNLKQIYTRSLIAAFC FHAYQQVSYMILARVSPVTHSVGNCVKRVVVIVTSVLFFRTPVSPINSLGTGVALAGV FLYSQLKRLKPKPKTA"

gene complement(24740. .25376)
/gene="OJ1381_H04.8"

mRNA complement(join(<24740. .24897,25172. .>25376))
/gene="OJ1381_H04.8"
/note="start and end point are not identified"

CDS complement(join(24740. .24897,25172. .25376))
/gene="OJ1381_H04.8"
/note="predicted by GeneMark.hmm etc."
/codon_start=1
/product="hypothetical protein"

```
                     /protein_id="BAD32979.1"
                     /db_xref="GI:50725510"
                     /translation="MEVSRSGPDPNSPLCQEATDSLLRRGDAERENEELDPEPPCGHC
                     TREMLRGKGMRKAMTTRPRRLAEVVTIATALDLNLCSGGGVGESGAVGATMEEEEMVV
                     EEVWGIGDKVDWLRSVFS"
     gene            complement(26622. .27293)
                     /gene="OJ1381_H04.9"
     mRNA            complement(join(<26622. .26706,26948. .27094,27220. .>27293))
                     /gene="OJ1381_H04.9"
                     /note="start and end point are not identified"
     CDS             complement(join(26622. .26706,26948. .27094,27220. .27293))
                     /gene="OJ1381_H04.9"
                     /note="predicted by GeneMark.hmm etc."
                     /codon_start=1
                     /product="hypothetical protein"

Alignment Scores:
Pred. No.:                 8.31e-13      Length:          130961
Score:                     576.00        Matches:         257
Percent Similarity:        31.28%        Conservative:    99
Best Local Similarity:     22.58%        Mismatches:      312
Query Match:               12.00%        Indels:          470
DB:                        8             Gaps:            41

US-09-920-705-3 (1-901) x AP004011 (1-130961)

Qy       144 LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPhe 163
Db     95385 CTGGACCTGAAATATTTGAATTTACAGGATAAAGGTTCAAGTAGCTCTATGGATGACAGA 95326

Qy       164 LeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsnGly 183
Db     95325 TACTCACAACATGCTGATCAAAGACGT--------------TTTACTGAACGAGGAAGG 95281

Qy       184 AsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspValGln 203
Db     95280 AATCAGCAATCAAGTCCTGAAAAATCAACTGATAAATCCAAAAGAAGCAGGCCAGCAGAA 95221

Qy       204 ProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGln 223
Db     95220 CCTAGTGAAGTATTATGGATAGGTTTTCCTGTTGGTCTGAAGGTA------GATGAGGCA 95167

Qy       224 IleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSer 243
Db     95166 ACTCTCTGGGAAGCCTTTTCACCTTTTGGTGAGGTTGTCAAGATAACTACATTCCCAGGG 95107

Qy       244 ArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluGly 263
Db     95106 CGTACTTATGCATTTGTCCAGTACACTACTATTGCAGCGGCATGCAGGGCGAAGGAAACA 95047

Qy       264 LeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSer------------ 279
Db     95046 CTGCAGGGAAATATTTTCAATAACCCTCGAGTTAGCATTTGCTTTTCTCGGAGTGACAGT 94987

Qy       280 ----------------------------------------------------AsnAsp 281
Db     94986 GTTTCAGCAGAATTTGGAAAAGGTTCCTTAGATGCCCCATATTCCCCCCATTTAAACTCT 94927

Qy       282 GluLeuProPro---------GluGlnAspAspThrSerPheTyrSerGlyMetLysArg 298
Db     94926 AGTGTTAGACCTATATTCAGGGAGCAAGATTTTGAAGATTTT----------------- 94885

Qy       299 SerArgThrAspMetPheAsnAsnAspPro------SerCysValSerSerProHisSer 316
Db     94884 CCTAGGGCTAGGCCTTTTGATAGTCCTCCAAGAGATATGTACATGCCATCTCCACATTAT 94825

Qy       317 ThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGly 336
Db     94824 ------------GGCCCTAAGAGACTTTCTAGA------------------------- 94804

Qy       337 AlaGluTyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGly 356
Db     94803 ---GATCATGATGATGTGGGTTTCAGCAGGGATAATTATTTGCGA-------------- 94762

Qy       357 ThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerPro------------ 372
Db     94761 --------------------TATGGACCTGGAGTAGAGCCTGATCCTAGATCTAATTTT 94723

Qy       373 ------------AlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrpGlu 388
Db     94722 GAACCTTTTAGGATACAAGGGCTCGGTCCAGAAAGAAGGATGTCTGAGGACCCATATGAA 94663

Qy       389 GlyTyr---------------------------------------------------- 390
Db     94662 CAGCATAGGCGTAGCCCTGCTGGTGATGCACCATGGCACAACATTCCATTCGAGCGATCT 94603

Qy       391 -----------------AspProAlaGlnLeuValArgGlu---------------Ser 399
Db     94602 CAGGGAGCCTTACCATTAGAGGATTCTCGGTATGCTAGGGAAGATCCATACCCATTTTCA 94543

Qy       400 LysArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGluArg 419
Db     94542 AAGAAGTTGAGGACTGGTGAAGCACATGACTCTGAACTTCCT---GAATACCCTTTCTCT 94486

Qy       420 SerPheGlyArgGlySerValAlaAla------------ArgProIleArgGlyProPro 435
Db     94485 GAATTTGATCGAGGGAAGGTTGGCTCTGCCTACCCAAGGAGGCCCTTCTATGGTGTGCCA 94426

Qy       436 AspSerAsp------------------------------------------------- 438
Db     94425 GATGATGACATACACCCCAGAGGCTATCAACTTGCTCCTATGCATGGTAGAAATCATGTT 94366

Qy       439 ----------------------------------HisIle------------------ 440
Db     94365 GATCCTTTAAGGAATCCAACTCCACTTGTAGATAGGCATATACCAGGGCATGCACAGGAC 94306

Qy       440 ------------------------------------------------------------ 440
Db     94305 AGCTTTTCTAGGCATGTAGAAGTGGAAAGATCAACTCCTGAATACCATGAACCCCTTCTC 94246

Qy       441 --------------TrpArgGlyMetIleAlaLysGlyGlyThrProValCysCysAla 455
Db     94245 AAGGAAGAATGGAAATGGGATGGTACAATAGCAAAGGGAGGCACACCAATTTGCCGAGCG 94186

Qy       456 ArgCysValProMetGlyLysGlyIleGluThrLysLeu---------------------- 468
Db     94185 CGATGCTTCCCTGTTGGGAAGGTTCTTAACTTCATGCTGTAAGTCTGGGTTCTTCACTGT 94126

Qy       468 ------------------------------------------------------------ 468
Db     94125 ATTTTCTTGTTCATTCTGAGTATCATTATTGACATCTACAACATTAGCTCTGCTTATCGT 94066

Qy       468 ------------------------------------------------------------ 468
Db     94065 CTTAATAAAAATCACAAAATTTCCTTCATGGTTCAAGATATATAATACATAGCAAATATA 94006

Qy       468 ------------------------------------------------------------ 468
Db     94005 TCATAGATATCGTTTTGCCTTTTACATCTAATTTATACTGCATGCTTTTGTTTGCACCTG 93946

Qy       468 ------------------------------------------------------------ 468
Db     93945 ATGAAGGCTTTTGTAGAGCTATGTTTTGTTCACGTTGTTCTTGATTTAATTATATCTCTG 93886

Qy       468 ------------------------------------------------------------ 468
Db     93885 CACCATTTGCCTCTACTCATAATCTTAATAAATACAGACGTTTGCCTTCACGGTCCAAAC 93826

Qy       468 ------------------------------------------------------------ 468
Db     93825 TACATAATCTAGTTGATCAATTTCTATCTGCATTCTTTTGGTTGTATCTGATGAGATGCC 93766

Qy       468 ------------------------------------------------------------ 468
Db     93765 TTTGGTGTTTATGCTCTGTTCTATTGTTCTAGATTTGAGCATCTCTAAGTTCAACATTTA 93706
```

```
Qy        469 ------------------------------ProGluValValAsnCysSerAlaArg 477
                                             ||||||   ::::::|||:::||||||
Db      93705 GTTTTAAACCGCTGTTGATTTCATACTTTCAGGCCCGAATTTTTGGATTGCACTGCTAGG 93646

Qy        478 ThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGluIleValPhe 497
              |||   |||   ||||||:::|||||||||       |||              :::||||||
Db      93645 ACAAGCCTGGAGATGCTCTCTAAGCACTATTACCAAGCTGCCAGCAGCTGGGTGGTGTTT 93586

Qy        498 PheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSer 517
              |||||||||:::   :::   |||   |||:::|||   ||||||:::   ||||||
Db      93585 TTTGTTCCAGAAAATGATGCTGACATGGCAGCCTATAATGAATTCATGAATTACCTTGGT 93526

Qy        518 SerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValPro 537
                 |||   ||||||   |||   ||||||   :::   ::::::|||||||||||||||
Db      93525 GATAAGCAGCGTGCAGCAGTTTGTAAACTTGGAGAAAGGAGCAGCTTATTTCTTGTTCCA 93466

Qy        538 ProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyVal 557
              ||||||||||||          ||||||:::|||      :::      :::   ||||||
Db      93465 CCCTCAGACTTCTCTGAACAAGTACTGAGGGTTCCAGGTAAAGTCAGCATATCTGGAGTC 93406

Qy        558 ValLeuLysLeuProProProAlaValProValThrAlaSerTyrArgGlnGluSer--- 576
              :::||||||                  |||::::::      |||:::
Db      93405 ATTCTGAAGTTTGAGCAGTCAGATCCAGAAGTTTCCTCGCCAACTCGCAAACCAGAAACA 93346

Qy        577 ------------------------------------------------------Gln 577
                                                                     :::
Db      93345 TTTGTGAGTCATTTGAACCATGATGTTCGTGCTCATGAGGATCTAGATGCATTGAGAAGA 93286

Qy        578 SerAsnPro------LeuHisTyrMetAspGlnAlaArgAsp------------SerPro 591
              ||||||      :::     :::   |||     |||              ||||||
Db      93285 ATCAACCCACCAGATATCAGGCCACTTCCTCAGGGTTCAGATTATCTCGGGTTGTCGCCT 93226

Qy        592 AlaAsn---------AlaSerHisSerLeuTyrProProArgGluAsnTyrIleArgGly 608
              :::         ||||||      |||   ||||||   :::
Db      93225 GGAAGCTATAATCCAGCAAGTGCACATTTGGTTCCGCCTTACAAG------TTTGGAAAT 93172

Qy        609 AlaProGluHisLeuThrAla------------------------------------ 615
              ||||||   :::|||   :::
Db      93171 GCTCCTTCATATCTAGAATCTGAATTAGCTCATCAAAAGCATCCACCTGACTCCCACAGG 93112

Qy        616 ------AlaSerLysProSerValSerGluProLeuArgIleProAsn---------- 629
                    |||                  |||      :::|||:::
Db      93111 GAGATAGCACATGACAAGCAGCAGCAACACCCAGATGTATTGCCCTCAAGATGGTCAGAT 93052

Qy        630 ---------AsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAlaThr 646
                    |||   :::|||   :::|||                  |||
Db      93051 AACATTTACAATCCAAGTCCAGGTTCTGGA--------------AATTTGAATTATTTG 93007

Qy        647 LeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHisGln------ 664
                 |||   :::|||   ||||||             ||||||::::::
Db      93006 GCTGAGAGTGCGATCCCACATACATCAACTGATAGGACACCAGAGGCATACTCATTTGCT 92947

Qy        665 ProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsn 684
              |||         |||         |||||||||                        :::
Db      92946 CCTCAAGGAGTACCAAAA------GTGAGTACA-----------------------TCA 92917

Qy        685 GlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHisAspAlaSerAsn 704
              |||   ||||||                                  :::||||||||||
Db      92916 GGGTATGCACCA--------------------------GTTGCAGATGAGGCATCAAAC 92884

Qy        705 GlnSerPhe------GlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeuProPro 722
                 |||:::          |||      ::|||                |||   ||||||
Db      92883 ATGTCCTACCCTCCCATGCAACCTGCATCACAGCAGGTAGTTAGACCTCAACAACCTCCA 92824

Qy        723 Pro---ProSerArgTyrProPro------------------------------Ala 730
                    |||         ||||||
Db      92823 TCTCTCCCATTATCGCTTCCACCAGAGCAACTTGCACAATTGGCCACTCTTCTTGCACAA 92764

Qy        731 SerAsnAsnPro----------------------AsnTyrThrSerGlyMetVal--- 741
              |||   |||                        |||      ||||||   :::
Db      92763 CAAAACCAACCAGGAAAAGAGCCTGTGGACAGCTTAAACAAAGAATCAGGATTCATACGG 92704

Qy        742 ------HisGlyAsn---------MetGlnTyrGlnSerGlnSerValAsnMetProGln 756
                       ||||||:::         |||   :::   |||   |||:::      |||
Db      92703 AATCCACATGGACATTCTTCAATGATGCCACACAGCTCAGGTTCTATC------CCTGTC 92650

Qy        757 LeuSerProLeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSer 776
                 :::   ||||||          |||                          :::|||
Db      92649 CAAAACTCATTGCCACCTGCTCCG----------------------CCATCTGCATCA 92614

Qy        777 AsnHisProValSerGlnProMetValGln--------------------------Gln 787
                       |||      |||   ||||||
Db      92613 CAGTTACAGGTTCATGCGCCACCAGTTCAAGGTTCAGTGCCACCAAATCCATCCATTATG 92554

Qy        788 TyrGlnProGluAlaSerMetProAsnGlnAsnTyrGlyProIleProSerTyrGlnGln 807
                 :::   |||   |||   ||||||:::   |||       |||:::|||
Db      92553 CATACACCGAATGCTCCTATGCCTTCTCACAACACTTTGCCTTTACCCCCTATGCATCCT 92494

Qy        808 AlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGln 827
                 :::                  |||   |||   :::       |||   :::
Db      92493 TCGGGA-----------------AATCCAGCTCATTCTTCCATGCCTTTGAGATCGTTC 92452

Qy        828 AlaAlaMetGlnProProAlaAspLysAlaAsnLeuGluProGlnAsnGlnAlaLeuArg 847
                       |||                  :::       :::   ||||||:::
Db      92451 GTCCCTCCTCTTCCTGAAGGCCCTCCACCCCTTAGGCAGCACACATCAAGTGCTCTACAA 92392

Qy        848 LeuGlnProMetIleSerGlyAspGlyGln-----------------GlyThrThrAsp 861
                 ||||||   :::            |||                        |||
Db      92391 GCGCAACCTGCACTTCCCTCTGGACCACAGACTAGCCAGCAACCGTCTGCTCAGGAAGAC 92332

Qy        862 GlyGluValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsnLeuLeu 881
                       |||      :::|||   |||:::||||||||   ||||||   ||||||
Db      92331 CATCACGGAGACCCTCAAAAGCGCCTTCAAGCAACATTGCAATTGGCAGCAACCCTACTT 92272

Qy        882 LeuGlnIleGlnGlnLysGlnGlnGlnGlnSerSerGlyThrProAlaGlyGln 899
                 ||||||||||||:::   :::                |||   ||||||
Db      92271 CAGCAGATACAGCAACAATCTAAA-----------------CCTGGTGGCCAG 92236
```

```
RESULT 15
ATT1P17
LOCUS       ATT1P17               137519 bp    DNA     linear   PLN 23-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone T1P17 (ESSA
            project).
ACCESSION   AL049730
VERSION     AL049730.2  GI:5823567
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 137519)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de
COMMENT     On Sep 5, 1999 this sequence version replaced gi:4725940.
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .137519
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     misc_feature    1. .18855
                     /note="overlap to BAC T4C9; please refer to this entry for
                     analysis and annotation"
     gene            17536. .20463
                     /gene="AT4g12420"
     gene            complement(join(17536. .17635,17716. .17915,18006. .18084,
                     18165. .18530,18612. .18978,19061. .19199,19285. .19555,
                     19663. .19772,20332. .20463))
```

```
                /gene="AT4g12420"
CDS             complement(join(17536. .17635,17716. .17915,18006. .18084,
                18165. .18530,18612. .18978,19061. .19199,19285. .19555,
                19663. .19772,20332. .20463))
                /gene="AT4g12420"
                /note="EST GB:Z37701 marks 3'end
                strong similarity to pollen-specific protein precursor -
                common tobacco, PIR2:S22495
                contains EST gb:N38363, R30492"
                /codon_start=1
                /product="putative pollen-specific protein"
                /protein_id="CAB41712.1"
                /db_xref="GI:4725941"
                /db_xref="GOA:Q9SU40"
                /db_xref="UniProt/TrEMBL:Q9SU40"
                /translation="MDLFKILLLVFFVNISFCFAADPYSFYNFEVSYITASPLGVPQQ
                VIAINGKFPGPTINVTTNENLVVNVRNKLDEGLLLHWNGIQQRRVSWQDGVLGTNCPI
                PPKWNWTYEFQVKDQIGSFFYFPSLHFQRASGGFGSFVVNPRAIIPVPFSTPDGDITV
                TIGDWYIRNHTALRKALDDGKDLGMPDGVLINGKGPYRYNDTLVADGIDFETITVHPG
                KTYRLRVSNVGISTSLNFRIQGHNLVLAESEGSYTVQQNYTSLDIHVGQSYSFLVTMD
                QNASSDYYIVASARVVNETIWRRVTGVGILKYTNSKGKAKGQLPPGPQDEFDKTFSMN
                QARSIRWNVSASGARPNPQGSFKYGSINVTDVYVLRNMPPVTISGKRRTTLNGISFKN
                PSTPIRLADKLKVKDVYKLDFPPKRPLTGPAKVATSIINGTYRGFMEVVLQNNDTKMQS
                YHMSGYAFFVVGMDYGEWTENSRGTYNKWDGIARSTIQVYPGAWSAILISLDNPGAWN
                LRTENLDSWYLGQETYVRVVNPDENNKTEFGHPDNVLYCGALSKLQKPQKVSSSASKS
                IGFTSLSMVVMALVMMMMLQH"
exon            complement(17536. .17635)
                /gene="AT4g12420"
                /number=1
intron          complement(17636. .17715)
                /gene="AT4g12420"
                /number=1
exon            complement(17716. .17915)
                /gene="AT4g12420"
                /number=2
intron          complement(17916. .18005)
                /gene="AT4g12420"
                /number=2
exon            complement(18006. .18084)
                /gene="AT4g12420"
                /number=3
intron          complement(18085. .18164)
                /gene="AT4g12420"
                /number=3
exon            complement(18165. .18530)
                /gene="AT4g12420"
                /number=4
intron          complement(18531. .18611)
                /gene="AT4g12420"
                /number=4
exon            complement(18612. .18978)
                /gene="AT4g12420"
                /number=5
intron          complement(18979. .19060)
                /gene="AT4g12420"
                /number=5
exon            complement(19061. .19199)
                /gene="AT4g12420"
                /number=6
intron          complement(19200. .19284)
                /gene="AT4g12420"
                /number=6
exon            complement(19285. .19555)
                /gene="AT4g12420"
                /number=7
intron          complement(19556. .19662)
                /gene="AT4g12420"
                /number=7
exon            complement(19663. .19772)
                /gene="AT4g12420"
                /number=8
intron          complement(19773. .20331)
                /gene="AT4g12420"
                /number=8
exon            complement(20332. .20463)
                /gene="AT4g12420"
                /number=9
misc_feature    25039. .26211
                /note="homology to fragments of putative reverse
                transcriptase"
repeat_region   27024. .32516
misc_feature    27260. .27596
                /note="EST GB:F19914 match, no obvious ORF"
repeat_region   28015. .28048
misc_feature    28046. .28260
                /note="EST GB:F19913 match, no obvious ORF"
misc_feature    31521. .31895
                /note="homology to fragments of putative reverse
                transcriptase"
gene            33075. .34941
                /gene="AT4g12430"
gene            complement(join(33075. .33140,33223. .33315,33414. .33484,
                33571. .33640,33722. .33799,33880. .34038,34112. .34219,
                34295. .34366,34469. .34609,34693. .34941))
                /gene="AT4g12430"
CDS             complement(join(33075. .33140,33223. .33315,33414. .33484,
                33571. .33640,33722. .33799,33880. .34038,34112. .34219,
                34295. .34366,34469. .34609,34693. .34941))
                /gene="AT4g12430"
                /note="strong similarity to trehalose-6-phosphate
                phosphatase - Arabidopsis thaliana, PID:g2944178
                contains EST gb:AA586291"
                /codon_start=1
                /product="putative trehalose-6-phosphate phosphatase
                (AtTPPA)"
                /protein_id="CAB41713.1"
                /db_xref="GI:4725942"
                /db_xref="GOA:Q9SU39"
                /db_xref="UniProt/TrEMBL:Q9SU39"
                /translation="MDLNSNHKSSVLKDPSPSVNQSRLGVSSRFMMSQWKKPAKLDDV
                RSNGWLDAMISSSPPRKKLVKDFNVEVAPEDDFAQRAWMVKYPSAISSFAHIAAQAKK
                KKIAVFLDYDGTLSPIVDDPDRAIMSDAMRSAVKDVASYFPTAIISGRSRDKVYQLVG
                LTELYYAGSHGMDIMTSSDGPNCFKSTDQQGKEVNLFQPAREFIPVIDEVFRTLVEKM
                KDIKGAKVENHKFCASVHYRNVDEKDWPIIAQRVHDHLKQYPRLRLTHGRKVLEVRPV
                IDWNKGRAVEFLLESLGLSNKDDLLPIYIGDDTTDEDAFKVLRDGNRGFGILVSSIPK
                ESNAFYSLRDPSEVKKFLKTLVKWAKLEKNSTGF"
exon            complement(33075. .33140)
                /gene="AT4g12430"
                /number=1
intron          complement(33141. .33222)
                /gene="AT4g12430"
                /number=1
exon            complement(33223. .33315)
                /gene="AT4g12430"
                /number=2
intron          complement(33316. .33413)
                /gene="AT4g12430"
                /number=2
exon            complement(33414. .33484)
                /gene="AT4g12430"
                /number=3
intron          complement(33485. .33570)
                /gene="AT4g12430"
                /number=3
exon            complement(33571. .33640)
                /gene="AT4g12430"
                /number=4
intron          complement(33641. .33721)
                /gene="AT4g12430"
                /number=4
exon            complement(33722. .33799)
                /gene="AT4g12430"
                /number=5
intron          complement(33800. .33879)
                /gene="AT4g12430"
                /number=5
```

```
     exon            complement(33880. .34038)
                     /gene="AT4g12430"
                     /number=6
     intron          complement(34039. .34111)
                     /gene="AT4g12430"
                     /number=6
     exon            complement(34112. .34219)
                     /gene="AT4g12430"
                     /number=7
     intron          complement(34220. .34294)
                     /gene="AT4g12430"
                     /number=7
     exon            complement(34295. .34366)
                     /gene="AT4g12430"
                     /number=8
     intron          complement(34367. .34468)
                     /gene="AT4g12430"
                     /number=8

Alignment Scores:
Pred. No.:              1.1e-12         Length:         137519
Score:                  573.50          Matches:        240
Percent Similarity:     35.37%          Conservative:   90
Best Local Similarity:  25.72%          Mismatches:     274
Query Match:            11.95%          Indels:         330
DB:                     8               Gaps:           36

US-09-920-705-3 (1-901) x ATT1P17 (1-137519)

Qy        145 GlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeu 164
Db     132829 CAGGAAAAGTCATCAACTGGATCACGCACTGAT-----------------GATATATAT 132870

Qy        165 ArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsnGlyAsn 184
Db     132871 CGCCATGATGAACAGCGGTCTGCAGCAAGAGGGTCGTCTTTTGTCCAAAGGGACTCCAGA 132930

Qy        185 MetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLys---------GlyAsp 201
Db     132931 ATGCGTTAT------GAGAGCCCAGACACATATAGCAAATCAAAAATGAATGATAGAAAT 132984

Qy        202 ValGlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAsp 221
Db     132985 GCAGAACCCAGTGAGGTACTATATATAGGGTTTCCT------GCTTCGTTGAAAGTAGAT 133038

Qy        222 GluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyr 241
Db     133039 GATGCGCTCTTGAGGAACGTCTTTTCTTCGTTTGGAGAAATAACAAAGGTCACGGTATTC 133098

Qy        242 ProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLys 261
Db     133099 CCTGGTCGCAGTTATGCATTTGTTCAATTCCGAAATCTGATGGCAGCTTGTAAGGCGAAA 133158

Qy        262 GluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAsp 281
Db     133159 GAAAGTCTTCAGGGAAAGTTATTTGGCAATCCTCGAGTGCATATTTGTTTTGCAAAGAGT 133218

Qy        282 Glu-----------------------------------------------LeuProProGlu 286
Db     133219 GAACCTTCCTCATCTGGCAGTGGAAGGGGTCCGTCGGGTAGATCACTCTCTCCACCTTAC 133278

Qy        287 Gln------AspAspThrSerPheTyrSerGlyMetLysArgSerArgThr---AspMet 303
Db     133279 AGATCTGTTGATCGACTGGGATCTTCAGAAGGTTATCTTCAGGATAGAAACTATGGAAGC 133338

Qy        304 PheAsnAsnAspProSerCysValSerSerProHis----------------------- 315
Db     133339 ATCAGCAGAATTCCCAGT---GTGAGGGAACCACATTACATAGAAGATAGGGATTTAGAA 133395

Qy        316 ----------------------------------------SerThrGlyIleProGly 321
Db     133396 GATTCTGAAGGTTACATTTTTAACAGGAAACGAGACTCAAGTAGCGATGGAGGTCCTGCA 133455

Qy        322 SerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAsp 341
Db     133456 TATGGAAGGTCG---AGATCTACGCATAGATTTCCTCAGGATATGCATGAATATCATGGT 133512

Qy        342 -ValValGlyLysGluProAsnTrp--------------------ArgArgProSerAl 354
Db     133513 AGTCCTGGGGAAA---------TGGGTACTTCATTCCGTGATAATCCTCACAGATTCCAG 133563

Qy        354 aAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGl 374
Db     133564 ACAAGGA---------GTTCAGAGTATGAAGAGCCATGGGATCTACCCGAAGATGACTAC 133614

Qy        374 nGlyThrArg------ArgProMetArgSerAsnProAspSerTrpGluGlyTyrAspPr 392
Db     133615 TACTACCAGGAAATCAAGAGATTGAAGACAAGATCCTCGCAGCCTGAGAGGCAGCTC-CC 133673

Qy        392 oAla---GlnLeu------ValArgGluSerLysArgThrArgArgAspGlySerValAs 409
Db     133674 AGGGCATCAGCTTTCTGGTATAGAGCAAGAAAGACGCCCCTTTTCAAGGGCATCTGCTGA 133733

Qy        409 pGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGly------------- 424
Db     133734 T---TTTTCCCCAAAAGATGCATTTGAGCGGAACTATGAGGCTGGACAGCTAAGATACAA 133790

Qy        424 ------------------------------------------------------------ 424
Db     133791 CCAAACAGTTGAGCAGCCGTTAAACCTGGCTATAAGGAATGGGGACAAGAGTAGCTTACG 133850

Qy        425 ----------------------------------------SerValAlaAlaArgProIl 431
Db     133851 AGAACCACATGAAGAGTTAATGGGAGGTTATCCTCTGCCATCAGTTGTCCCTGAAAGAAA 133910

Qy        431 eArgGlyProProAsp-----------------SerAspHisIleTrpArgGlyMetIl 445
Db     133911 AAGGTACACCCCTGAACTGAATCGACCATCACTAAAGGACTGGAACTGGGAAGGGACTAT 133970

Qy        445 eAlaLysGlyGlyThrProValCysCysAlaArgCysValProMetGlyLysGlyIleGl 465
Db     133971 TGCTAAGGGAGGCAATCCCATTTGTCGAGCAAAATGCTTTCCTGTGGGCAAAGTGATGGA 134030

Qy        465 uThrLysLeu------------------------------------------------- 468
Db     134031 CATGATGCTGTAAGTAGCTCTAAGTCTTTATTTTTCTGTATTTCTTGTTTTGATTGGCTC 134090

Qy        468 ------------------------------------------------------------ 468
Db     134091 TCTATATTTGGCTTAAGTTCATTCAACCAAGTATAAGTCAAATCCGTTTGTGTATTGGTT 134150

Qy        468 ------------------------------------------------------------ 468
Db     134151 TAAGACTTTTGCTTAATTTCATTGTTCTTTTTGTGTATTGCTTTTAGACTTCTGATCAAA 134210

Qy        469 ----------------------------------ProGluValValAsnCysSerAl 476
Db     134211 TTAATTTTTATTGGTTTGTTATGGTTATAAATGCAGGCCTGAGTTTCTAGATTGCACGGC 134270

Qy        476 aArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGluIleVa 496
Db     134271 AAGAACTGGTTTAGACATGCTGGCTAAGCATTACTACCAATCATCTAAAGCATGGGTGGT 134330

Qy        496 lPhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLe 516
Db     134331 TTTCTTTGTTCCTGGAAGTGATGCTGATATCGTGTTTTATGATGAGTTTATGCATTATCT 134390

Qy        516 uSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuVa 536
Db     134391 GGAGGAGAAGCAACGGGCAGCTGTTTCTAAATTGGATGACACAACAACGTTGTTTCTGGT 134450

Qy        536 lProProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGl 556
Db     134451 GCCTCCATCTGATTTCTCCGAGAAAGTACTTAAAGTTCCTGGGAAACTAAGCATCTCTGG 134510

Qy        556 yValValLeuLysLeuProProProAlaValProValThrAlaSerTyrArgGlnGluSe 576
```

```
                ||||:::|||:::|||
Db     134511 AGTTATTCTACGTTTA---------------------------------------- 134526

Qy        576 rGlnSerAsnProLeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHi 596
                                                                    |||
Db     134527 ----------------------------------------GAATGTGGTGGTTCTGG 134543

Qy        596 sSerLeuTyrProProArgGluAsnTyrIleArgGlyAlaProGluHisLeuThrAlaAl 616
               |||      |||   ::::::         |||         :::   ||||||
Db     134544 CTCCGGTGGGCCTGTTCAACAGCAAGGTGAGAGAAAAGATACAGACTTGCTAACTTATTA 134603

Qy        616 aSerLysProSerValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGl 636
                   :::   |||   ||||||          :::   |||   |||      ||
Db     134604 TGGTGAAACATCATATTCAGAA-----------GCTAGTGGAGCGTTTCCTGATGTGGG 134651

Qy        636 yValSerLeuThrProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGl 656
                |
Db     134652 A------------------------------------------------------AA 134654

Qy        656 nProAlaAlaProGluSerHisGlnProMetSerGlyProSerThrValValSerThrAl 676
               |||      |||                      ||||||:::      :::   :::||
Db     134655 CCCACGTATTCCG--------------------GGTCCAACAGCTTTTCTGCGTAGTGC 134693

Qy        676 aHisGlnSerAsnGlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyPr 696
                |   :::   |||                      |||:::              ||
Db     134694 GGGTCGTGACAAT-------------------------CAAAGTGCATCAATGGATCC 134726

Qy        696 oGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrPr 716
                |   |||   |||      :::|||         :::|||
Db     134727 ATAC---GTCGAGGACAAGCATGATCAGTTGAGTCACCGGTATTCTGGA----------- 134772

Qy        716 oAlaGlyGlnLeuProProProProSerArgTyrProProAlaSerAsnAsnProAsnTy 736
                                  |||   :::||||||         ||||||
Db     134773 -------------------------TCGGACTGGCCGCCTCGTGATACAAACCCCAGAAG 134807

Qy        736 rThr-----------------------------SerGlyMetValHisGlyAsnMetGl 746
                :::                           ||||||   |||   |||      ||
Db     134808 TTCACCCTTCATAGATCATACTGTTCAAAAACACAGCGGGTTTGTTCCTGGGAAACAGCA 134867

Qy        746 nTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeuProAsnMetProHisAs 766
                |                                                    :::
Db     134868 A----------------------------------------------------AATGC 134873

Qy        766 nAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProValSerGlnProMetValGl 786
                :::   |||   |||                  |||           |||:::
Db     134874 AGATCTAAGTCGTTAT-----------------CATGATACGGAAACTCCAGTCCCAGC 134915

Qy        786 nGlnTyrGlnProGlu-----------------AlaSerMetProAsnGlnAsnTyrGl 800
                :::|||||||||                  :::|||:::|||   |||
Db     134916 AGGGTTTCAGCCTGAACAGCTTACACATTTAGCTTCCTCGCTACCCAGACAGCAA----- 134970

Qy        800 yProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAs 820
                      ||||||                  |||   ||||||   :::
Db     134971 ---------------CAACAGGTACAG---------AACACTCCAAATCAGCCTGAGAG 135005

Qy        820 nLeuAsnPro-----------------SerGlnPheGlnAlaAlaMetGlnProProAl 834
                |||                   |||      |||   |||   |||   |||::
Db     135006 ATATGCACCAGAAGGCCGAGCAAGTTTCAGCCATTTGCAACATGCG---CAGACACCAAG 135062

Qy        834 aAspLysAlaAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGl 854
                :          :::   ||||||||||||   :::::::::|||          |||
Db     135063 CATACCTCAATTGGTAACTCCACAAAATCAAAATGTACAGATACAA---------AGTAG 135113

Qy        854 yAspGlyGlnGlyThrThrAspGlyGluValAspLysAsnGlnArgTyrGlnSerThrLe 874
                :::   |||       :::   |||   :::      :::|||   |||:::|||||
Db     135114 CAACAGCCAACAACAAGAAGAAACAGAGGCTAATCCACAGAAGCGTCTTCAAGCAACATT 135173

Qy        874 uGlnPheAlaAlaAsnLeuLeuLeuGlnIleGlnGln 886
                ||||   ||||||   ||||||   ||||||||||||
Db     135174 ACAGCTAGCAGCTGCACTTCTCCAGCAGATTCAACAA 135210

Search completed: April  8, 2005, 12:01:29
```

Job time : 10611 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2005, 07:37:17 ; Search time 1152 Seconds
(without alignments)
4629.937 Million cell updates/sec

Title: US-09-920-705-3
Perfect score: 4801
Sequence: 1 MALSMKPFRADDSGFQSNNL.....LQIQQKQQQQSSGTPAGQGP 901

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09920705/runat_07042005_101812_20618/app_query.fasta_1.1095
-DB=N_Geneseq -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09920705_@CGN_1_1_581_@runat_07042005_101812_20618 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4801	100.0	2706	6	ABA95167	Aba95167 Arabidops
	2	4656.5	97.0	3171	12	ADN73104	Adn73104 Thale cre
	3	4392	91.5	4593	6	ABA95166	Aba95166 Arabidops
c	4	1436	29.9	801	6	ABA95168	Aba95168 Arabidops
	5	651.5	13.6	3715	6	ABA95170	Aba95170 Arabidops
	6	466	9.7	349	6	ABA95169	Aba95169 Arabidops
	7	336.5	7.0	6836	8	ABZ21692	Abz21692 Human RBM
	8	302	6.3	2190	3	AAZ51263	Aaz51263 Human RNA
	9	302	6.3	3127	13	ACN37956	Acn37956 Tumour-as
	10	302	6.3	87980	8	AAD53223_4	Continuation (5 of
	11	297	6.2	2673	13	ADS95035	Ads95035 Human the
	12	297	6.2	2673	13	ADS95197	Ads95197 Human the
	13	287.5	6.0	3084	4	ABL05395	Abl05395 Drosophil
c	14	287.5	6.0	9744	4	ABL05394	Abl05394 Drosophil
	15	276.5	5.8	3383	8	ABZ21697	Abz21697 Human RBM
	16	273	5.7	3192	4	AAH17592	Aah17592 Human cDN
	17	273	5.7	3312	8	ABZ21695	Abz21695 Human RBM
	18	273	5.7	3423	8	ABZ21696	Abz21696 Human RBM
	19	271.5	5.7	10656	12	ADO00904	Ado00904 Mouse hom
	20	268	5.6	3228	6	ABZ11870	Abz11870 Human pol
	21	268	5.6	3228	12	ADM44388	Adm44388 Novel hum
	22	265	5.5	10995	12	ADO00903	Ado00903 Human hom
	23	265	5.5	11145	8	ACD13385	Acd13385 Human DNA
	24	265	5.5	11680	3	AAC76700	Aac76700 Human ORF
	25	265	5.5	12227	8	ACC50291	Acc50291 Breast ca
	26	265	5.5	12227	12	ADK60461	Adk60461 Angiogene
	27	265	5.5	12227	12	ADK60762	Adk60762 Angiogene
	28	265	5.5	12227	12	ADP73084	Adp73084 Angiogene
	29	265	5.5	12227	13	ADS88605	Ads88605 Human hou
	30	265	5.5	12227	13	ADP23327	Adp23327 PRO polyp
	31	261	5.4	2505	13	ADR07405	Adr07405 Full leng
	32	261	5.4	11295	5	ADL45263	Adl45263 Human ova
	33	254	5.3	438	6	ABL94207	Abl94207 Arabidops
	34	247	5.1	3046	4	ABL09397	Abl09397 Drosophil
	35	247	5.1	3389	4	ABL09385	Abl09385 Drosophil
	36	246.5	5.1	1436	10	ADF81563	Adf81563 Leukaemia
	37	246.5	5.1	1477	12	ADL13065	Adl13065 Human ste
	38	246.5	5.1	1871	4	AAS61035	Aas61035 Human can
	39	246	5.1	2228	13	ADR24549	Adr24549 Breast ca
	40	246	5.1	2482	12	ADQ23974	Adq23974 Human sof
	41	243	5.1	2228	2	AAQ34554	Aaq34554 cDNA enco
	42	240	5.0	1766	6	ABI99779	Abi99779 Mouse isc
	43	236.5	4.9	8147	6	ABS73313	Abs73313 DNA encod
	44	236.5	4.9	8147	10	ADK11302	Adk11302 Human CRE
	45	236.5	4.9	8677	12	ADQ85476	Adq85476 Human tum

ALIGNMENTS

RESULT 1
ABA95167
ID ABA95167 standard; cDNA; 2706 BP.
XX
AC ABA95167;
XX
DT 20-MAY-2002 (first entry)
XX
DE Arabidopsis floral induction gene (FPA) cDNA sequence.
XX
KW FPA; floral induction; photoperiod; plant; flowering; FLC; FRI; frigida;
KW flowering locus C; transgenic; gene; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 1. .2706
FT /*tag= a
FT /product= "FPA protein"
FT misc_binding 46. .279
FT /*tag= b
FT /note= "RNA binding region"
FT misc_binding 283. .522
FT /*tag= c
FT /note= "RNA binding region"
FT misc_binding 610. .852
FT /*tag= d
FT /note= "RNA binding region"
XX

PN WO200212518-A2.
XX
PD 14-FEB-2002.
XX
PF 02-AUG-2001; 2001WO-US024427.
XX
PR 03-AUG-2000; 2000US-0222550P.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Amasino RM, Schomburg FM, Michaels SD, Patton D;
XX
DR WPI; 2002-227160/28.
DR P-PSDB; ABB07659.
XX
PT Novel isolated DNA sequence comprising coding sequence for floral
PT induction gene (FPA) gene which controls flowering time in plants, useful
PT for altering flowering time of plants and for downregulating flowering
PT locus C (FLC) mRNA activity.
XX
PS Claim 5; Page 29-34; 39pp; English.
XX
CC The invention relates to the FPA gene (floral induction promoter in
CC plants during both long and short day photoperiods) from Arabidopsis
CC thaliana. The FPA coding sequence is useful for altering flowering time
CC in a plant, and for down-regulating flowering locus C (FLC) mRNA
CC activity. The FPA gene can accelerate flowering in several late-flowering
CC mutant backgrounds and can fully compensate for addition of two naturally
CC occurring genes, frigida (FRI) and FLC which confer late-flowering
CC phenotypes. Over expression of FPA can compensate for the delaying effect
CC caused by short days on floral induction, and decreases FLC mRNA in
CC plants containing the FLC gene. Fragments of the FPA gene can also act to
CC decrease activity of an endogenous FPA gene by modifying the expression
CC of the endogenous FPA gene and expression of a portion of polypeptide
CC encoded by FPA gene can also lead to a delay of flowering in a plant. The
CC present sequence represents the A. thaliana FPA protein encoding cDNA
XX
SQ Sequence 2706 BP; 779 A; 628 C; 617 G; 682 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3.74e-246	Length:	2706
Score:	4801.00	Matches:	901
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-920-705-3 (1-901) x ABA95167 (1-2706)

```
Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 60

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 120

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 180

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 240

Qy         81 SerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTCAAATTAAGATCGAATACGCACGACCGGCAAAACCTTGTAAGAGTCTATGGGTGGGT 300

Qy        101 GlyIleGlyProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAATCGGCCCTAATGTCTCCAAGGATGACCTGGAGGAAGAGTTCAGCAAGTTTGGGAAA 360

Qy        121 IleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMet 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATCGAGGATTTTAGGTTTCTCAGAGAACGCAAGACAGCTTTCATTGATTATTATGAGATG 420

Qy        141 AspAspAlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArg 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATGATGCTTTACAGGCTAAGAGCATGAATGGAAAGCCTATGGGTGGTAGCTTTTTGCGT 480

Qy        161 ValAspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsn 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTTGATTTTCTCCGGTCACAAGCGCCAAAAAAAGAACAATGGGCTGGCTCTTACGATAAC 540

Qy        181 ArgAsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGly 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGAAATGGCAATATGAATCATAAACCGCAGTATCCTCACTCATATGAAGACTTTAAAGGA 600

Qy        201 AspValGlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GATGTCCAGCCAAGTAAGGTTCTGTGGATTGGGTTCCCTCCTACTGCTACACAATGCAAT 660

Qy        221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATGAGCAAATTCTGCACAATGCGATGATACTCTTTGGTGAGATCGAGAGGGTAAAAAGT 720

Qy        241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TACCCATCAAGGAATTTTGCACTTGTGGAGTTTAGGAGCGCGGAGGAAGCTCGCCAATGC 780

Qy        261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGGAAGGCCTACAGGGGAGGTTATTCAATAATCCTAGAATCAAAATTATGTACTCAAAC 840

Qy        281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GATGAGTTGCCTCCTGAGCAAGACGATACTAGTTTTTACTCTGGTATGAAACGGTCAAGG 900

Qy        301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACAGATATGTTCAATAATGATCCTTCATGTGTATCTTCTCCTCATTCTACTGGAATTCCT 960

Qy        321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsn 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGGTCTATGAGGCCCCTCAGAGGTACGAATGAGCGTTCATATAATGGTGCAGAATACAAT 1020

Qy        341 AspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GACGTTGTTGGTAAGGAGCCAAACTGGAGGAGGCCATCTGCAAATGGAACTGGAATACTC 1080

Qy        361 ProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMet 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCATCTCCAACAGGACCTGGAATCCTCCCATCTCCTGCACAAGGTACGAGGCGCCCTATG 1140

Qy        381 ArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuValArgGluSerLys 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AGGTCAAACCCCGATTCTTGGGAAGGATATGATCCTGCTCAGTTGGTCAGAGAAAGTAAA 1200

Qy        401 ArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGluArgSer 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGAACCAGAAGAGATGGATCAGTGGACGGTTTTACTCCAATGGGTGTCGATGAGAGGTCA 1260

Qy        421 PheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAspSerAspHisIle 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTTGGTCGAGGTTCAGTTGCTGCTAGACCTATCCGTGGCCCCCCTGATTCTGATCACATA 1320

Qy        441 TrpArgGlyMetIleAlaLysGlyGlyThrProValCysCysAlaArgCysValProMet 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGAGAGGAATGATTGCCAAGGGTGGAACTCCCGTCTGTTGTGCTCGTTGTGTACCTATG 1380

Qy        461 GlyLysGlyIleGluThrLysLeuProGluValValAsnCysSerAlaArgThrAspLeu 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GGAAAGGGGATTGAAACTAAACTGCCTGAGGTCGTCAATTGTTCAGCAAGAACTGATTTG 1440

Qy        481 AsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGluIleValPhePheValPro 500
```

```
Db       1441 AATATGCTCGCTAAACATTACGCCGTTGCCATTGGATGTGAGATCGTTTTTTTCGTACCA 1500

Qy        501 AspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAsp 520

Db       1501 GACAGGGAAGAAGATTTTGCGTCTTACACTGAATTTCTCCGGTACCTTAGCTCAAAAGAT 1560

Qy        521 ArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProProSerAsp 540

Db       1561 CGGGCGGGTGTTGCCAAATTAGATGATGGTACAACTTTATTCTTGGTGCCTCCATCAGAT 1620

Qy        541 PheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLys 560

Db       1621 TTCTTAACTGATGTACTCCAAGTGACCCGTCAAGAACGGCTATATGGTGTTGTTCTCAAG 1680

Qy        561 LeuProProProAlaValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnPro 580

Db       1681 TTACCCCCGCCAGCCGTTCCTGTTACAGCATCATACAGACAAGAATCTCAGTCCAATCCT 1740

Qy        581 LeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrPro 600

Db       1741 CTGCATTATATGGATCAAGCCCGGGATTCACCTGCCAATGCTAGTCACAGTTTATATCCT 1800

Qy        601 ProArgGluAsnTyrIleArgGlyAlaProGluHisLeuThrAlaAlaSerLysProSer 620

Db       1801 CCTAGGGAAAATTACATTAGGGGTGCACCAGAACATTTGACAGCTGCTTCAAAACCATCT 1860

Qy        621 ValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThr 640

Db       1861 GTTAGCGAGCCTCTCAGAATACCTAATAATGCAGCGCCTCAAGCTGGGGTTAGTTTAACT 1920

Qy        641 ProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro 660

Db       1921 CCGGAGCTTTTAGCCACTCTGGCATCTATTCTCCCTGCAACTTCTCAACCTGCTGCCCCT 1980

Qy        661 GluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsn 680

Db       1981 GAGAGTCACCAACCTATGTCAGGACCTTCAACAGTTGTTTCCACAGCACATCAGTCCAAT 2040

Qy        681 GlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHis 700

Db       2041 GGACTGTACAATGGAGAAGCACCGTCTCAAGCTTGGAAAAGAGGTCCACAAACAGTTCAT 2100

Qy        701 AspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeu 720

Db       2101 GATGCGTCAAATCAGTCATTCCAACAATACGGAAATCAGTACACTCCAGCTGGGCAACTA 2160

Qy        721 ProProProProSerArgTyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMet 740

Db       2161 CCTCCTCCTCCTTCGCGTTACCCTCCAGCTTCAAACAACCCCAACTACACTAGTGGAATG 2220

Qy        741 ValHisGlyAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeu 760

Db       2221 GTCCATGGCAACATGCAATACCAGAGCCAATCTGTTAACATGCCTCAGCTGTCTCCGTTA 2280

Qy        761 ProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProVal 780

Db       2281 CCAAATATGCCTCATAATAATTATTCCATGTACACTCAGGGTTCGTCAAATCATCCTGTT 2340

Qy        781 SerGlnProMetValGlnGlnTyrGlnProGluAlaSerMetProAsnGlnAsnTyrGly 800

Db       2341 TCTCAGCCCATGGTCCAGCAATACCAACCAGAAGCGTCCATGCCAAACCAAAACTATGGT 2400

Qy        801 ProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsn 820

Db       2401 CCAATTCCAAGTTATCAGCAAGCTAATTTTCATGGCGTAACAACAAATCAGGCACAGAAC 2460

Qy        821 LeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGlu 840

Db       2461 TTAAACCCTTCCCAATTTCAAGCTGCCATGCAACCACCAGCAGATAAGGCAAATTTAGAG 2520

Qy        841 ProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThr 860

Db       2521 CCACAAAACCAAGCACTACGATTGCAGCCTATGATCTCTGGGGATGGTCAGGGTACAACA 2580

Qy        861 AspGlyGluValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsnLeu 880

Db       2581 GATGGGGAGGTCGATAAGAATCAGAGATACCAGTCAACACTACAATTTGCAGCAAACCTT 2640

Qy        881 LeuLeuGlnIleGlnGlnLysGlnGlnGlnGlnSerSerGlyThrProAlaGlyGlnGly 900

Db       2641 CTTCTCCAGATACAGCAGAAACAGCAGCAACAGTCTTCAGGTACTCCGGCTGGACAGGGG 2700

Qy        901 Pro 901
               |||
Db       2701 CCT 2703
```

RESULT 2
ADN73104
ID ADN73104 standard; cDNA; 3171 BP.
XX
AC ADN73104;
XX
DT 15-JUL-2004 (first entry)
XX
DE Thale cress cDNA upregulated in E2Fa/Dpa expressing plants SeqID 999.
XX
KW gene; ss; plant; transgenic; E2Fa/DPa transcription factor;
KW growth regulator; animal feed product; thale cress;
KW cell wall biosynthesis; nitrogen metabolism; carbon metabolism.
XX
OS Arabidopsis thaliana.
XX
PN WO2004035798-A2.
XX
PD 29-APR-2004.
XX
PF 20-OCT-2003; 2003WO-EP011658.
XX
PR 18-OCT-2002; 2002EP-00079408.
XX
PA (CROP-) CROPDESIGN NV.
XX
PI Inze D, De Veylder L, Vlieghe K;
XX
DR WPI; 2004-348466/32.
DR P-PSDB; ADN73105.
XX
PT Altering plant characteristics, useful for producing plants for enzyme or
PT pharmaceutical production comprises modifying in a plant, expression of
PT one or more nucleic acids and/or modifying level or activity of one or
PT more proteins.
XX
PS Claim 1; SEQ ID NO 999; 134pp; English.
XX
CC This invention relates to a novel method for altering one or more plant
CC characteristics. Specifically, it refers to identifying genes that are up
CC - or down-regulated in transgenic plants overexpressing the heterodimeric
CC E2Fa/DPa transcription factor of Arabidopsis and using these sequences to
CC alter plant characteristics accordingly. The present invention describes
CC generating transgenic plants for the production of growth regulators,
CC enzymes, therapeutics, pharmaceuticals and animal feed products, where
CC the altered plant characteristics are selected from increased yield or
CC biomass, enhanced survival capacity, stress tolerance, plant architecture
CC or physiology, altered endoreduplication, biochemistry, signal
CC transduction, storage lipid mobilisation and/or altered photosynthesis,
CC each relative to the corresponding wild type plants. Accordingly, these
CC sequences can also be useful as positive or negative selectable markers
CC during transformation of cells or tissues. The identified genes play a
CC role in a variety of biological processes such as DNA replication, cell
CC wall biosynthesis, nitrogen and/ or carbon metabolism or they function as
CC transcription factors. This polynucleotide sequence is thale cress cDNA
CC upregulated 1.3 fold or more in plants overexpressing the E2Fa/DPa
CC transcription factor, given in an exemplification of the invention.
XX
SQ Sequence 3171 BP; 910 A; 723 C; 709 G; 829 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.13e-238	Length:	3171
Score:	4656.50	Matches:	892
Percent Similarity:	85.04%	Conservative:	6
Best Local Similarity:	84.47%	Mismatches:	3
Query Match:	96.99%	Indels:	155
DB:	12	Gaps:	3

US-09-920-705-3 (1-901) x ADN73104 (1-3171)

```
Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu  20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT  60

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg  40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 120

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 180

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly  80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 240

Qy         81 Ser------------------------------------------------------------  81
              :::
Db        241 ACGAATGAGCTCTGCAAAAGAATACATCAAACACATGTTTACTCTCGTTTATGCGAAGAA 300

Qy         81 ------------------------------------------------------------  81

Db        301 GGTTTAATCCGATTGTCGTTTATCTGCAAACTTGTATCCGCGTTTAGCAGTCTGATTTCA 360

Qy         81 ------------------------------------------------------------  81

Db        361 GGTTCGCTTCAGATGTTAAATCTCACAAGCTTGAGTATGAATGTATTGCGTACTCCACTC 420

Qy         82 ----------------------------GlnIleLysIleGluTyr---------------  87
                                          :::|||:::::::::
Db        421 GAAATTCGCAATGATAAGAAATTTTGGAGTTTAAAAGTTCAGTTTCCTGCAAAAATCAAA 480

Qy         87 ------------------------------------------------------------  87

Db        481 CCGCGTGGAGAAATTTGTCTATGGCGGTACGTAGATATCAATAACTGCATCTGCGAAACA 540

Qy         87 ------------------------------------------------------------  87

Db        541 GAAAATTCTGGCAACCGTATCGTCCTAACTCCTAGTATCGCTGGCACATATCCATATCAG 600

Qy         87 ------------------------------------------------------------  87

Db        601 TGTGAGGGAGTTTTGGGCTTGTTTTGGCTCGATGCTTCAGGAAGACAAATTACGTGGTTG 660

Qy         87 ------------------------------------------------------------  87

Db        661 TTAAGGCGGCTAACTCTACCAATCAGAAACGCTTTATTCGAAGAACCATGTTTGTTCCTC 720

Qy         88 ---AlaArgPro---AlaLysProCysLysSerLeuTrpValGlyGlyIleGlyProAsn 105
                 :::   |||   ||||||||||||||||||||||||||||||||||||||||||||
Db        721 AATTCCCATCCCTACGCAAAACCTTGTAAGAGTCTATGGGTGGGTGGAATCGGCCCTAAT 780

Qy        106 ValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArg 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTCTCCAAGGATGACCTGGAGGAAGAGTTCAGCAAGTTTGGGAAAATCGAGGATTTTAGG 840

Qy        126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTCTCAGAGAACGCAAGACAGCTTTCATTGATTATTATGAGATGGATGATGCTTTACAG 900

Qy        146 AlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeuArg 165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        901 GCTAAGAGCATGAATGGAAAGCCTATGGGTGGTAGCTTTTTGCGTGTTGATTTTCTCCGG 960

Qy        166 SerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsnGlyAsnMet 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TCACAAGCGCCAAAAAAAGAACAATGGGCTGGCTCTTACGATAACAGAAATGGCAATATG 1020

Qy        186 AsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspValGlnProSer 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AATCATAAACCGCAGTATCCTCACTCATATGAAGACTTTAAAGGAGATGTCCAGCCAAGT 1080

Qy        206 LysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeu 225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AAGGTTCTGTGGATTGGGTTCCCTCCTACTGCTACACAATGCAATGATGAGCAAATTCTG 1140

Qy        226 HisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsn 245
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CACAATGCGATGATACTCTTTGGTGAGATCGAGAGGGTAAAAAGTTACCCATCAAGGAAT 1200

Qy        246 PheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeuGln 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTTGCACTTGTGGAGTTTAGGAGCGCGGAGGAAGCTCGCCAATGCAAGGAAGGCCTACAG 1260

Qy        266 GlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProPro 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GGGAGGTTATTCAATAATCCTAGAATCAAAATTATGTACTCAAACGATGAGTTGCCTCCT 1320

Qy        286 GluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsn 305
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GAGCAAGACGATACTAGTTTTTACTCTGGTATGAAACGGTCAAGGACAGATATGTTCAAT 1380

Qy        306 AsnAspProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArgPro 325
              ||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AATGATCCTTCATTTGTATCTTCTCCTCATTCTACTGGAATTCCTGGGTCTATGAGGCCC 1440

Qy        326 LeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValValGlyLys 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CTCAGAGGTACAAATGAGCGTTCATATAATGGTGCAGAATACAATGACGTTGTTGGTAAG 1500

Qy        346 GluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGly 365
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GAGCCAAACTGGAGGAGGCCATCTGCAAATGGAACTGGAATACTCCCATCTCCAACAGGA 1560

Qy        366 ProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAsp 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTGGAATCCTCCCATCTCCTGCACAAGGTACGAGGCGCCCTATGAGGTCAAACCCCGAT 1620

Qy        386 SerTrpGluGlyTyrAspProAlaGlnLeuValArgGluSerLysArgThrArgArgAsp 405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TCTTGGGAAGGATATGATCCTGCTCAGTTGGTCAGAGAAAGTAAACGAACCAGAAGAGAT 1680

Qy        406 GlySerValAspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGlySer 425
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GGATCAGTGGACGGTTTTACTCCAATGGGTGTCGATGAGAGGTCATTTGGTCGAGGTTCA 1740

Qy        426 ValAlaAlaArgProIleArgGlyProProAspSerAspHisIleTrpArgGlyMetIle 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GTTGCTGCTAGACCTATCCGTGGCCCCCCTGATTCTGATCACATATGGAGAGGAATGATT 1800

Qy        446 AlaLysGlyGlyThrProValCysCysAlaArgCysValProMetGlyLysGlyIleGlu 465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GCCAAGGGTGGAACTCCCGTCTGTTGTGCTCGTTGTGTACCTATGGGAAAGGGGATTGAA 1860

Qy        466 ThrLysLeuProGluValValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLys 485
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACTAAACTGCCTGAGGTCGTCAATTGTTCAGCAAGAACTGATTTGAATATGCTCGCTAAA 1920

Qy        486 HisTyrAlaValAlaIleGlyCysGluIleValPhePheValProAspArgGluGluAsp 505
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CATTACGCCGTTGCCATTGGATGTGAGATCGTTTTTTTCGTACCAGACAGGGAAGAAGAT 1980

Qy        506 PheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAla 525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TTTGCGTCTTACACTGAATTTCTCCGGTACCTTAGCTCAAAAGATCGGGCGGGTGTTGCC 2040
```

```
Qy       526 LysLeuAspAspGlyThrThrLeuPheLeuValProProSerAspPheLeuThrAspVal 545
Db      2041 AAATTAGATGATGGTACAACTTTATTCTTGGTGCCTCCATCAGATTTCTTAACTGATGTA 2100

Qy       546 LeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLysLeuProProProAla 565
Db      2101 CTCCAAGTGACCCGTCAAGAACGGCTATATGGTGTTGTTCTCAAGTTACCCCCGCCAGCC 2160

Qy       566 ValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnProLeuHisTyrMetAsp 585
Db      2161 GTTCCTGTTACAGCATCATACAGACAAGAATCTCAGTCCAATCCTCTGCATTATATGGAT 2220

Qy       586 GlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGluAsnTyr 605
Db      2221 CAAGCCCGGGATTCACCTGCCAATGCTAGTCACAGTTTATATCCTCCTAGGGAAAATTAC 2280

Qy       606 IleArgGlyAlaProGluHisLeuThrAlaAlaSerLysProSerValSerGluProLeu 625
Db      2281 ATTAGGGGTGCACCAGAACATTTGACAGCTGCTTCAAAACCATCTGTTAGCGAGCCTCTC 2340

Qy       626 ArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAla 645
Db      2341 AGAATACCTAATAATGCAGCGCCTCAAGCTGGGGTTAGTTTAACTCCGGAGCTTTTAGCC 2400

Qy       646 ThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHisGlnPro 665
Db      2401 ACTCTGGCATCTATTCTCCCTGCAACTTCTCAACCTGCTGCCCCTGAGAGTCACCAACCT 2460

Qy       666 MetSerGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGly 685
Db      2461 ATGTCAGGACCTTCAACAGTTGTTTCCACAGCACATCAGTCCAATGGACTGTACAATGGA 2520

Qy       686 GluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHisAspAlaSerAsnGln 705
Db      2521 GAAGCACCGTCTCAAGCTTGGAAAAGAGGTCCACAAACAGTTCATGATGCGTCAAATCAG 2580

Qy       706 SerPheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeuProProProProSer 725
Db      2581 TCATTCCAACAATACGGAAATCAGTACACTCCAGCTGGGCAACTACCTCCTCCTCCTTCG 2640

Qy       726 ArgTyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMet 745
Db      2641 CGTTACCCTCCAGCTTCAAACAACCCCAACTACACTAGTGGAATGGTCCATGGCAACATG 2700

Qy       746 GlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeuProAsnMetProHis 765
Db      2701 CAATACCAGAGCCAATCTGTTAACATGCCTCAGCTGTCTCCGTTACCAAATATGCCTCAT 2760

Qy       766 AsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProValSerGlnProMetVal 785
Db      2761 AATAATTATTCCATGTACACTCAGGGTTCGTCAAATCATCCTGTTTCTCAGCCCATGGTC 2820

Qy       786 GlnGlnTyrGlnProGluAlaSerMetProAsnGlnAsnTyrGlyProIleProSerTyr 805
Db      2821 CAGCAATACCAACCAGAAGCGTCCATGCCAAACCAAAACTATGGTCCAATTCCAAGTTAT 2880

Qy       806 GlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGln 825
Db      2881 CAGCAAGCTAATTTTCATGGCGTAACAACAAATCAGGCACAGAACTTAAACCCTTCCCAA 2940

Qy       826 PheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGluProGlnAsnGlnAla 845
Db      2941 TTTCAAGCTGCCATGCAACCACCAGCAGATAAGGCAAATTTAGAGCCACAAAACCAAGCA 3000

Qy       846 LeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThrAspGlyGluValAsp 865
Db      3001 CTACGATTGCAGCCTATGATCTCTGGGGATGGTCAGGGTACAACAGATGGGGAGGTCGAT 3060

Qy       866 LysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsnLeuLeuLeuGlnIleGln 885
Db      3061 AAGAATCAGAGATACCAGTCAACACTACAATTTGCAGCAAACCTTCTTCTCCAGATACAG 3120

Qy       886 GlnLysGlnGlnGlnGlnSerSerGlyThrProAlaGlyGlnGlyPro 901
Db      3121 CAGAAACAGCAGCAACAGTCTTCAGGTACTCCGGCTGGACAGGGGCCT 3168
```

RESULT 3
ABA95166
ID ABA95166 standard; DNA; 4593 BP.
XX
AC ABA95166;
XX
DT 20-MAY-2002 (first entry)
XX
DE Arabidopsis floral induction gene (FPA) sequence.
XX
KW FPA; floral induction; photoperiod; plant; flowering; FLC; FRI; frigida;
KW flowering locus C; transgenic; gene; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200212518-A2.
XX
PD 14-FEB-2002.
XX
PF 02-AUG-2001; 2001WO-US024427.
XX
PR 03-AUG-2000; 2000US-0222550P.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Amasino RM, Schomburg FM, Michaels SD, Patton D;
XX
DR WPI; 2002-227160/28.
DR P-PSDB; ABB07659.
XX
PT Novel isolated DNA sequence comprising coding sequence for floral
PT induction gene (FPA) gene which controls flowering time in plants, useful
PT for altering flowering time of plants and for downregulating flowering
PT locus C (FLC) mRNA activity.
XX
PS Claim 5; Page 27-29; 39pp; English.
XX
CC The invention relates to the FPA gene (floral induction promoter in
CC plants during both long and short day photoperiods) from Arabidopsis
CC thaliana. The FPA coding sequence is useful for altering flowering time
CC in a plant, and for down-regulating flowering locus C (FLC) mRNA
CC activity. The FPA gene can accelerate flowering in several late-flowering
CC mutant backgrounds and can fully compensate for addition of two naturally
CC occurring genes, frigida (FRI) and FLC which confer late-flowering
CC phenotypes. Over expression of FPA can compensate for the delaying effect
CC caused by short days on floral induction, and decreases FLC mRNA in
CC plants containing the FLC gene. Fragments of the FPA gene can also act to
CC decrease activity of an endogenous FPA gene by modifying the expression
CC of the endogenous FPA gene and expression of a portion of polypeptide
CC encoded by FPA gene can also lead to a delay of flowering in a plant. The
CC present sequence represents the A. thaliana FPA gene sequence
XX
SQ Sequence 4593 BP; 1275 A; 958 C; 978 G; 1382 T; 0 U; 0 Other;

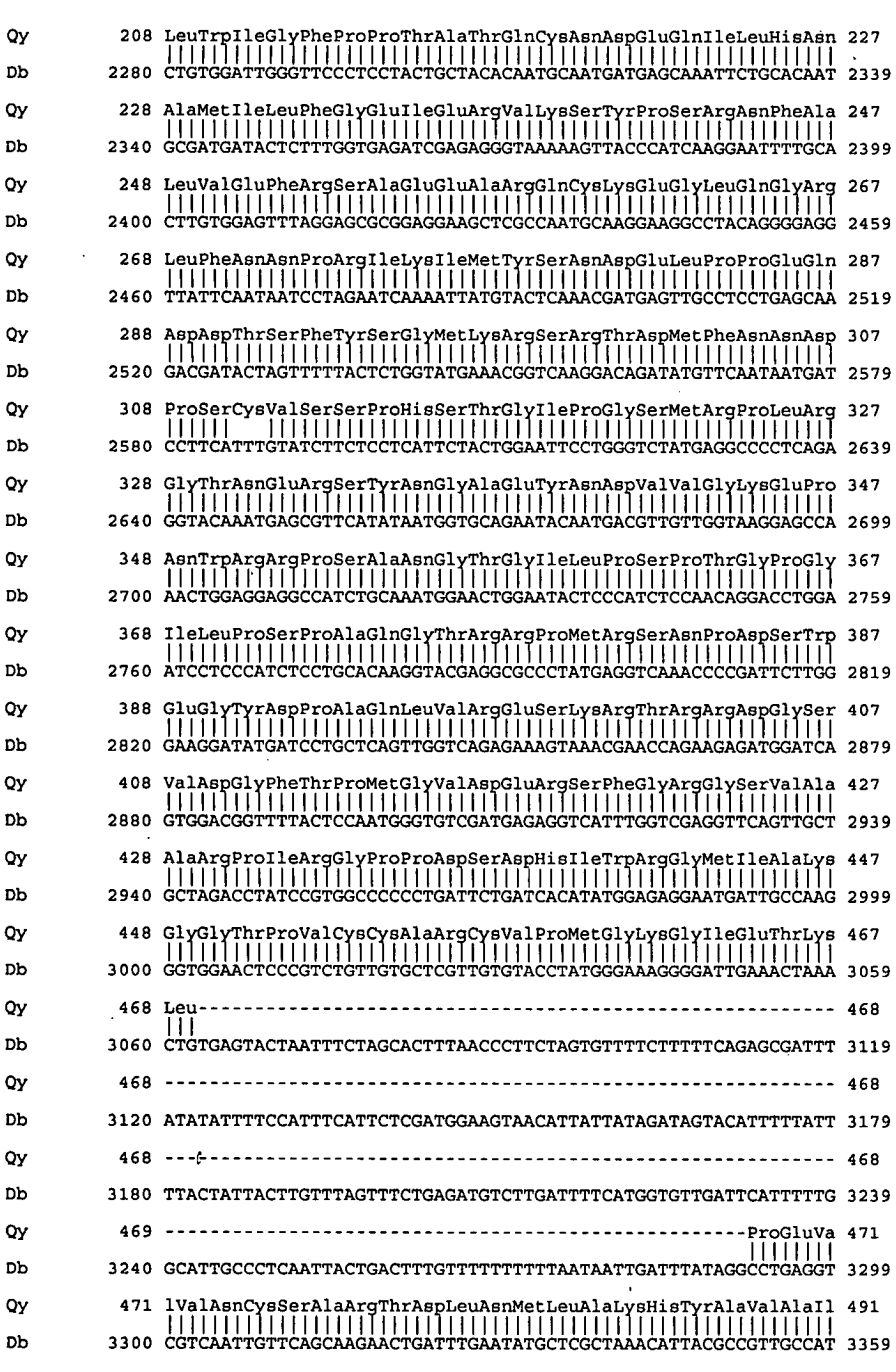
Alignment Scores:

Pred. No.:	3.64e-224	Length:	4593
Score:	4392.00	Matches:	899
Percent Similarity:	58.72%	Conservative:	0
Best Local Similarity:	58.72%	Mismatches:	2
Query Match:	91.48%	Indels:	631
DB:	6	Gaps:	4

US-09-920-705-3 (1-901) x ABA95166 (1-4593)

```
Qy         1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
Db         1 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 60

Qy        21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
```

```
Db        61 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 120
Qy        41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
Db       121 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 180
Qy        61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
Db       181 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 240
Qy        81 SerGlnIleLysIleGluTyrAlaArgPro---------------------------- 90
Db       241 AGTCAAATTAAGATCGAATACGCACGACCGGTTTGTTCTTATCTATATCTTCGTTTGTTC 300
Qy        90 ---------------------------------------------------------- 90
Db       301 TCTAACTTTGATTGTCTTTTGTCAACGATTATACTCTTTTTGCGAATTCATAGTCCAGGT 360
Qy        90 ---------------------------------------------------------- 90
Db       361 TCACAAAACTTTGATGATGCTTGTTTAGTCCAAAAATTTCTTGTTGAATCTGTTTTTTTT 420
Qy        90 ---------------------------------------------------------- 90
Db       421 TTCCTCATCGTACAAATCAAAGTCGAAACCTAGTTTTTTTCTATTATACGTCGTTAGCTT 480
Qy        90 ---------------------------------------------------------- 90
Db       481 AAGGCGAAACCTGATCCGATCGAAACGTCTTTTCTCAAATTACTTTGGTTATATCGAACT 540
Qy        90 ---------------------------------------------------------- 90
Db       541 CGCGCAAAGCCAAACCACAGAGAAGCTCTGCAAAATTTGATGTTAAAGCATATATAACTC 600
Qy        90 ---------------------------------------------------------- 90
Db       601 TTAGCGAATGAGCTCTGCAAAAGAATACATCAAACACATGTTTACTCTCGTTTATGCGAA 660
Qy        90 ---------------------------------------------------------- 90
Db       661 GAAGGTTTAATCCGATTGTCGTTTATCTGCAAACTTGTATCCGCGTTTAGCAGTCTGATT 720
Qy        90 ---------------------------------------------------------- 90
Db       721 TCAGGTTCGCTTCAGATGTTAAATCTCACAAGCTTGAGTATGAATGTATTGCGTACTCCA 780
Qy        90 ---------------------------------------------------------- 90
Db       781 CTCGAAATTCGCAATGGTGAAAATTTGGAAGCGAGCAAACTCTATCATCGGCCAAACAGA 840
Qy        90 ---------------------------------------------------------- 90
Db       841 TAAGAAATTTTGGAGTTTAAAAGTTCAGTTTCCTGCAAAAATCAAACCGCGTGGAGAAAT 900
Qy        90 ---------------------------------------------------------- 90
Db       901 TTGTCTATGGCGGTACGTAGATATCAATAACTGCATCTGCGAAACAGAAAATTCTGGCAA 960
Qy        90 ---------------------------------------------------------- 90
Db       961 CCGTATCGTCCTAACTCCTAGTATCGCTGGCACATATCCATATCAGTGTGAGGGAGTTTT 1020
Qy        90 ---------------------------------------------------------- 90
Db      1021 GGGCTTGTTTTGGCTCGATGCTTCAGGAAGACAAATTACGTGGTTGTTAAGGCGGCTAAC 1080
Qy        90 ---------------------------------------------------------- 90
Db      1081 TCTACCAATCAGAAACGCTTTATTCGAAGAACCATGTTTGTTCCTCAATTCCCATCCCTA 1140
Qy        90 ---------------------------------------------------------- 90

Db      1141 CGTACAATCTGGGCTTTCCCATTGTAGTTCCTTAGGAAGTTGACTTCTTCACAACCATTC 1200
Qy        90 ---------------------------------------------------------- 90
Db      1201 TGGGATTGGTACAGTTGCAGTGGACAAGACAAATATTCATTTGCAGGCAGACTCAAATTT 1260
Qy        90 ---------------------------------------------------------- 90
Db      1261 CAATGTCTTGCCGGCTTTGAGTACTTAAATGGAGTGTTCAGGGATTGGTTTATTTGGGAC 1320
Qy        90 ---------------------------------------------------------- 90
Db      1321 TCAGGCAATGGATAAAGACAGGAATGTTTTGCAGAAAGTATTAATGTCTTTCCGGCTTTG 1380
Qy        90 ---------------------------------------------------------- 90
Db      1381 GGGACTACATACTTCTCCTACAGACAAAATCTGATGTTTAGGCAAGAGAAACTATTAACA 1440
Qy        90 ---------------------------------------------------------- 90
Db      1441 CTGATTTAATAGAGAAAGAGGAGATGGTTTCTCCTTGCGGCAATTTTATTTGTTTAGGAA 1500
Qy        90 ---------------------------------------------------------- 90
Db      1501 AGCAATTGATATGAATTGGTGTCGTAGTGTAGTTGAAATTACTAGTTAGTTTGTGTGTTT 1560
Qy        90 ---------------------------------------------------------- 90
Db      1561 AGTTTCCTTGATGTTTGATGCTTTATTCTTGGCAACCTATCTGGGTAGTATCGCCTTCTT 1620
Qy        91 ---------------------AlaLysProCysLysSerLeuTrpValGlyGlyIleGly 103
Db      1621 ATCGACCTTTTCTTGTTGCAGGCAAAACCTTGTAAGAGTCTATGGGTGGGTGGAATCGGC 1680
Qy       104 ProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAsp 123
Db      1681 CCTAATGTCTCCAAGGATGACCTGGAGGAAGAGTTCAGCAAGTTTGGGAAAATCGAGGAT 1740
Qy       124 PheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAla 143
Db      1741 TTTAGGTTTCTCAGAGAACGCAAGACAGCTTTCATTGATTATTATGAGATGGATGATGCT 1800
Qy       144 LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPhe 163
Db      1801 TTACAGGCTAAGAGCATGAATGGAAAGCCTATGGGTGGTAGCTTTTTGCGTGTTGATTTT 1860
Qy       164 LeuArgSerGlnAlaProLysLys---------------------------------- 171
Db      1861 CTCCGGTCACAAGCGCCAAAAAAAGTAAGCACTCTTGTGGCATTTGATTTTTACTTTTGA 1920
Qy       171 ---------------------------------------------------------- 171
Db      1921 AAACGCTCCAGTAAACATTTTGTTTAGTTTCATAATTTGCGTCAAACTGATAGGGCTGAG 1980
Qy       171 ---------------------------------------------------------- 171
Db      1981 CTCTGTCTTGTGCCCCTAGGAGCAGTATTTACTCGTCTCTATTTCATTGTAGAGTAGGCT 2040
Qy       172 ----------------------------------------------GluGlnTrpAlaGly 176
Db      2041 CAACTTCTTAAGTCTGAAATCAAGTTACCTTTGTGTTATCTTCAGGAACAATGGGCTGGC 2100
Qy       177 SerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGln------------------ 190
Db      2101 TCTTACGATAACAGAAATGGCAATATGAATCATAAACCGCA-GGTTAGTCTTGAATGTTG 2159
Qy       190 ---------------------------------------------------------- 190
Db      2160 AAAGTATGTCTCTTGTTACTAGTGATATGTATAGGTTACTGGTTTTGACGTTTTGTTATA 2219
Qy       191 ---------TyrProHisSerTyrGluAspPheLysGlyAspValGlnProSerLysVal 207
Db      2220 TTCTTACAGTATCCTCACTCATATGAAGACTTTAAAGGAGATGTCCAGCCAAGTAAGGTT 2279
```


Qy 208 LeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHisAsn 227
Db 2280 CTGTGGATTGGGTTCCCTCCTACTGCTACACAATGCAATGATGAGCAAATTCTGCACAAT 2339
Qy 228 AlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsnPheAla 247
Db 2340 GCGATGATACTCTTTGGTGAGATCGAGAGGGTAAAAAGTTACCCATCAAGGAATTTTGCA 2399
Qy 248 LeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeuGlnGlyArg 267
Db 2400 CTTGTGGAGTTTAGGAGCGCGGAGGAAGCTCGCCAATGCAAGGAAGGCCTACAGGGGAGG 2459
Qy 268 LeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProProGluGln 287
Db 2460 TTATTCAATAATCCTAGAATCAAAATTATGTACTCAAACGATGAGTTGCCTCCTGAGCAA 2519
Qy 288 AspAspThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsnAsp 307
Db 2520 GACGATACTAGTTTTTACTCTGGTATGAAACGGTCAAGGACAGATATGTTCAATAATGAT 2579
Qy 308 ProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArgProLeuArg 327
Db 2580 CCTTCATTTGTATCTTCTCCTCATTCTACTGGAATTCCTGGGTCTATGAGGCCCCTCAGA 2639
Qy 328 GlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValValGlyLysGluPro 347
Db 2640 GGTACAAATGAGCGTTCATATAATGGTGCAGAATACAATGACGTTGTTGGTAAGGAGCCA 2699
Qy 348 AsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGly 367
Db 2700 AACTGGAGGAGGCCATCTGCAAATGGAACTGGAATACTCCCATCTCCAACAGGACCTGGA 2759
Qy 368 IleLeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrp 387
Db 2760 ATCCTCCCATCTCCTGCACAAGGTACGAGGCGCCCTATGAGGTCAAACCCCGATTCTTGG 2819
Qy 388 GluGlyTyrAspProAlaGlnLeuValArgGluSerLysArgThrArgArgAspGlySer 407
Db 2820 GAAGGATATGATCCTGCTCAGTTGGTCAGAGAAAGTAAACGAACCAGAAGAGATGGATCA 2879
Qy 408 ValAspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGlySerValAla 427
Db 2880 GTGGACGGTTTTACTCCAATGGGTGTCGATGAGAGGTCATTTGGTCGAGGTTCAGTTGCT 2939
Qy 428 AlaArgProIleArgGlyProProAspSerAspHisIleTrpArgGlyMetIleAlaLys 447
Db 2940 GCTAGACCTATCCGTGGCCCCCCTGATTCTGATCACATATGGAGAGGAATGATTGCCAAG 2999
Qy 448 GlyGlyThrProValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLys 467
Db 3000 GGTGGAACTCCCGTCTGTTGTGCTCGTTGTGTACCTATGGGAAAGGGGATTGAAACTAAA 3059
Qy 468 Leu---------- 468
Db 3060 CTGTGAGTACTAATTTCTAGCACTTTAACCCTTCTAGTGTTTTCTTTTTCAGAGCGATTT 3119
Qy 468 ---------- 468
Db 3120 ATATATTTTCCATTTCATTCTCGATGGAAGTAACATTATTATAGATAGTACATTTTTATT 3179
Qy 468 ---------- 468
Db 3180 TTACTATTACTTGTTTAGTTTCTGAGATGTCTTGATTTTCATGGTGTTGATTCATTTTTG 3239
Qy 469 ----------ProGluVa 471
Db 3240 GCATTGCCCTCAATTACTGACTTTGTTTTTTTTTTAATAATTGATTTATAGGCCTGAGGT 3299
Qy 471 lValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIl 491
Db 3300 CGTCAATTGTTCAGCAAGAACTGATTTGAATATGCTCGCTAAACATTACGCCGTTGCCAT 3359
Qy 491 eGlyCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGl 511
Db 3360 TGGATGTGAGATCGTTTTTTTCGTACCAGACAGGGAAGAAGATTTTGCGTCTTACACTGA 3419
Qy 511 uPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyTh 531
Db 3420 ATTTCTCCGGTACCTTAGCTCAAAAGATCGGGCGGGTGTTGCCAAATTAGATGATGGTAC 3479
Qy 531 rThrLeuPheLeuValProProSerAspPheLeuThrAspValLeuGlnValThrArgGl 551
Db 3480 AACTTTATTCTTGGTGCCTCCATCAGATTTCTTAACTGATGTACTCCAAGTGACCCGTCA 3539
Qy 551 nGluArgLeuTyrGlyValValLeuLysLeuProProProAlaValProValThrAlaSe 571
Db 3540 AGAACGGCTATATGGTGTTGTTCTCAAGTTACCCCCGCCAGCCGTTCCTGTTACAGCATC 3599
Qy 571 rTyrArgGlnGluSerGlnSerAsnProLeuHisTyrMetAspGlnAlaArgAspSerPr 591
Db 3600 ATACAGACAAGAATCTCAGTCCAATCCTCTGCATTATATGGATCAAGCCCGGGATTCACC 3659
Qy 591 oAlaAsnAlaSerHisSerLeuTyrProProArgGluAsnTyrIleArgGlyAlaProGl 611
Db 3660 TGCCAATGCTAGTCACAGTTTATATCCTCCTAGGGAAAATTACATTAGGGGTGCACCAGA 3719
Qy 611 uHisLeuThrAlaAlaSerLysProSerValSerGluProLeuArgIleProAsnAsnAl 631
Db 3720 ACATTTGACAGCTGCTTCAAAACCATCTGTTAGCGAGCCTCTCAGAATACCTAATAATGC 3779
Qy 631 aAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAlaThrLeuAlaSerIleLe 651
Db 3780 AGCGCCTCAAGCTGGGGTTAGTTTAACTCCGGAGCTTTTAGCCACTCTGGCATCTATTCT 3839
Qy 651 uProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMetSerGlyProSerTh 671
Db 3840 CCCTGCAACTTCTCAACCTGCTGCCCCTGAGAGTCACCAACCTATGTCAGGACCTTCAAC 3899
Qy 671 rValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyGluAlaProSerGlnAl 691
Db 3900 AGTTGTTTCCACAGCACATCAGTCCAATGGACTGTACAATGGAGAAGCACCGTCTCAAGC 3959
Qy 691 aTrpLysArgGlyProGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTyrGl 711
Db 3960 TTGGAAAAGAGGTCCACAAACAGTTCATGATGCGTCAAATCAGTCATTCCAACAATACGG 4019
Qy 711 yAsnGlnTyrThrProAlaGlyGlnLeuProProProProSerArgTyrProProAlaSe 731
Db 4020 AAATCAGTACACTCCAGCTGGGCAACTACCTCCTCCTCCTTCGCGTTACCCTCCAGCTTC 4079
Qy 731 rAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMetGlnTyrGlnSerGlnSe 751
Db 4080 AAACAACCCCAACTACACTAGTGGAATGGTCCATGGCAACATGCAATACCAGAGCCAATC 4139
Qy 751 rValAsnMetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyrSerMetTy 771
Db 4140 TGTTAACATGCCTCAGCTGTCTCCGTTACCAAATATGCCTCATAATAATTATTCCATGTA 4199
Qy 771 rThrGlnGlySerSerAsnHisProValSerGlnProMetValGlnGlnTyrGlnProGl 791
Db 4200 CACTCAGGGTTCGTCAAATCATCCTGTTTCTCAGCCCATGGTCCAGCAATACCAACCAGA 4259
Qy 791 uAlaSerMetProAsnGlnAsnTyrGlyProIleProSerTyrGlnGlnAlaAsnPheHi 811
Db 4260 AGCGTCCATGCCAAACCAAAACTATGGTCCAATTCCAAGTTATCAGCAAGCTAATTTTCA 4319
Qy 811 sGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGl 831
Db 4320 TGGCGTAACAACAAATCAGGCACAGAACTTAAACCCTTCCCAATTTCAAGCTGCCATGCA 4379
Qy 831 nProProAlaAspLysAlaAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMe 851
Db 4380 ACCACCAGCAGATAAGGCAAATTTAGAGCCACAAAACCAAGCACTACGATTGCAGCCTAT 4439
Qy 851 tIleSerGlyAspGlyGlnGlyThrThrAspGlyGluValAspLysAsnGlnArgTyrGl 871

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4440 GATCTCTGGGGATGGTCAGGGTACAACAGATGGGGAGGTCGATAAGAATCAGAGATACCA 4499

Qy        871 nSerThrLeuGlnPheAlaAlaAsnLeuLeuLeuGlnIleGlnGlnLysGlnGlnGlnGl 891
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4500 GTCAACACTACAATTTGCAGCAAACCTTCTTCTCCAGATACAGCAGAAACAGCAGCAACA 4559

Qy        891 nSerSerGlyThrProAlaGlyGlnGlyPro 901
                |||||||||||||||||||||||||||||
Db       4560 GTCTTCAGGTACTCCGGCTGGACAGGGGCCT 4590
```

RESULT 4
ABA95168/c
ID ABA95168 standard; DNA; 801 BP.
XX
AC ABA95168;
XX
DT 20-MAY-2002 (first entry)
XX
DE Arabidopsis floral induction gene (FPA) antisense fragment.
XX
KW FPA; floral induction; photoperiod; plant; flowering; FLC; FRI; frigida;
KW flowering locus C; transgenic; gene; antisense; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200212518-A2.
XX
PD 14-FEB-2002.
XX
PF 02-AUG-2001; 2001WO-US024427.
XX
PR 03-AUG-2000; 2000US-0222550P.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Amasino RM, Schomburg FM, Michaels SD, Patton D;
XX
DR WPI; 2002-227160/28.
XX
PT Novel isolated DNA sequence comprising coding sequence for floral
PT induction gene (FPA) gene which controls flowering time in plants, useful
PT for altering flowering time of plants and for downregulating flowering
PT locus C (FLC) mRNA activity.
XX
PS Claim 5; Page 37; 39pp; English.
XX
CC The invention relates to the FPA gene (floral induction promoter in
CC plants during both long and short day photoperiods) from Arabidopsis
CC thaliana. The FPA coding sequence is useful for altering flowering time
CC in a plant, and for down-regulating flowering locus C (FLC) mRNA
CC activity. The FPA gene can accelerate flowering in several late-flowering
CC mutant backgrounds and can fully compensate for addition of two naturally
CC occurring genes, frigida (FRI) and FLC which confer late-flowering
CC phenotypes. Over expression of FPA can compensate for the delaying effect
CC caused by short days on floral induction, and decreases FLC mRNA in
CC plants containing the FLC gene. Fragments of the FPA gene can also act to
CC decrease activity of an endogenous FPA gene by modifying the expression
CC of the endogenous FPA gene and expression of a portion of polypeptide
CC encoded by FPA gene can also lead to a delay of flowering in a plant. The
CC present sequence represents the A. thaliana FPA gene antisense fragment
XX
SQ Sequence 801 BP; 206 A; 207 C; 166 G; 222 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	7.51e-68	Length:	801
Score:	1436.00	Matches:	265
Percent Similarity:	99.62%	Conservative:	0
Best Local Similarity:	99.62%	Mismatches:	1
Query Match:	29.91%	Indels:	0
DB:	6	Gaps:	0

US-09-920-705-3 (1-901) x ABA95168 (1-801)

```
Qy        197 AspPheLysGlyAspValGlnProSerLysValLeuTrpIleGlyPheProProThrAla 216
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        799 GACTTTAAAGGAGATGTCCAGCCAAGTAAGGTTCTGTGGATTGGGTTCCCTCCTACTGCT 740

Qy        217 ThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGlu 236
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 ACACAATGCAATGATGAGCAAATTCTGCACAATGCGATGATACTCTTTGGTGAGATCGAG 680

Qy        237 ArgValLysSerTyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGlu 256
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 AGGGTAAAAAGTTACCCATCAAGGAATTTTGCACTTGTGGAGTTTAGGAGCGCGGAGGAA 620

Qy        257 AlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIle 276
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 GCTCGCCAATGCAAGGAAGGCCTACAGGGGAGGTTATTCAATAATCCTAGAATCAAAATT 560

Qy        277 MetTyrSerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMet 296
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 ATGTACTCAAACGATGAGTTGCCTCCTGAGCAAGACGATACTAGTTTTTACTCTGGTATG 500

Qy        297 LysArgSerArgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSer 316
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 AAACGGTCAAGGACAGATATGTTCAATAATGATCCTTCATTTGTATCTTCTCCTCATTCT 440

Qy        317 ThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGly 336
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 ACTGGAATTCCTGGGTCTATGAGGCCCCTCAGAGGTACAAATGAGCGTTCATATAATGGT 380

Qy        337 AlaGluTyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGly 356
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 GCAGAATACAATGACGTTGTTGGTAAGGAGCCAAACTGGAGGAGGCCATCTGCAAATGGA 320

Qy        357 ThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThr 376
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 ACTGGAATACTCCCATCTCCAACAGGACCTGGAATCCTCCCATCTCCTGCACAAGGTACG 260

Qy        377 ArgArgProMetArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuVal 396
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 AGGCGCCCTATGAGGTCAAACCCCGATTCTTGGGAAGGATATGATCCTGCTCAGTTGGTC 200

Qy        397 ArgGluSerLysArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyVal 416
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 AGAGAAAGTAAACGAACCAGAAGAGATGGATCAGTGGACGGTTTTACTCCAATGGGTGTC 140

Qy        417 AspGluArgSerPheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAsp 436
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 GATGAGAGGTCATTTGGTCGAGGTTCAGTTGCTGCTAGACCTATCCGTGGCCCCCCTGAT 80

Qy        437 SerAspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrProValCysCysAlaArg 456
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 TCTGATCACATATGGAGAGGAATGATTGCCAAGGGTGGAACTCCCGTCTGTTGTGCTCGT 20

Qy        457 CysValProMetGlyLys 462
                |||||||||||||||||
Db         19 TGTGTACCTATGGGAAAG 2
```

RESULT 5
ABA95170
ID ABA95170 standard; DNA; 3715 BP.
XX
AC ABA95170;
XX
DT 20-MAY-2002 (first entry)
XX
DE Arabidopsis floral induction gene (FPA) fragment.
XX
KW FPA; floral induction; photoperiod; plant; flowering; FLC; FRI; frigida;
KW flowering locus C; transgenic; gene; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200212518-A2.

XX
PD 14-FEB-2002.
XX
PF 02-AUG-2001; 2001WO-US024427.
XX
PR 03-AUG-2000; 2000US-0222550P.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Amasino RM, Schomburg FM, Michaels SD, Patton D;
XX
DR WPI; 2002-227160/28.
XX
PT Novel isolated DNA sequence comprising coding sequence for floral
PT induction gene (FPA) gene which controls flowering time in plants, useful
PT for altering flowering time of plants and for downregulating flowering
PT locus C (FLC) mRNA activity.
XX
PS Claim 5; Page 38; 39pp; English.
XX
CC The invention relates to the FPA gene (floral induction promoter in
CC plants during both long and short day photoperiods) from Arabidopsis
CC thaliana. The FPA coding sequence is useful for altering flowering time
CC in a plant, and for down-regulating flowering locus C (FLC) mRNA
CC activity. The FPA gene can accelerate flowering in several late-flowering
CC mutant backgrounds and can fully compensate for addition of two naturally
CC occurring genes, frigida (FRI) and FLC which confer late-flowering
CC phenotypes. Over expression of FPA can compensate for the delaying effect
CC caused by short days on floral induction, and decreases FLC mRNA in
CC plants containing the FLC gene. Fragments of the FPA gene can also act to
CC decrease activity of an endogenous FPA gene by modifying the expression
CC of the endogenous FPA gene and expression of a portion of polypeptide
CC encoded by FPA gene can also lead to a delay of flowering in a plant. The
CC present sequence represents the A. thaliana FPA gene fragment comprising
CC the FPA promoter and intron
XX
SQ Sequence 3715 BP; 1055 A; 728 C; 711 G; 1221 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.97e-25	Length:	3715
Score:	651.50	Matches:	171
Percent Similarity:	27.23%	Conservative:	0
Best Local Similarity:	27.23%	Mismatches:	0
Query Match:	13.57%	Indels:	457
DB:	6	Gaps:	1

US-09-920-705-3 (1-901) x ABA95170 (1-3715)

```
Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1832 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 1891

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1892 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 1951

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1952 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 2011

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2012 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 2071

Qy         81 SerGlnIleLysIleGluTyrAlaArgPro---------------------------- 90
              ||||||||||||||||||||||||||||
Db       2072 AGTCAAATTAAGATCGAATACGCACGACCGGTTTGTTCTTATCTATATCTTCGTTTGTTC 2131

Qy         90 ------------------------------------------------------------ 90

Db       2132 TCTAACTTTGATTGTCTTTTGTCAACGATTATACTCTTTTTGCGAATTCATAGTCCAGGT 2191

Qy         90 ------------------------------------------------------------ 90

Db       2192 TCACAAAACTTTGATGATGCTTGTTTAGTCCAAAAATTTCTTGTTGAATCTGTTTTTTTT 2251

Qy         90 ------------------------------------------------------------ 90

Db       2252 TTCCTCATCGTACAAATCAAAGTCGAAACCTAGTTTTTTTCTATTATACGTCGTTAGCTT 2311

Qy         90 ------------------------------------------------------------ 90

Db       2312 AAGGCGAAACCTGATCCGATCGAAACGTCTTTTCTCAAATTACTTTGGTTATATCGAACT 2371

Qy         90 ------------------------------------------------------------ 90

Db       2372 CGCGCAAAGCCAAACCACAGAGAAGCTCTGCAAAATTTGATGTTAAAGCATATATAACTC 2431

Qy         90 ------------------------------------------------------------ 90

Db       2432 TTAGCGAATGAGCTCTGCAAAAGAATACATCAAACACATGTTTACTCTCGTTTATGCGAA 2491

Qy         90 ------------------------------------------------------------ 90

Db       2492 GAAGGTTTAATCCGATTGTCGTTTATCTGCAAACTTGTATCCGCGTTTAGCAGTCTGATT 2551

Qy         90 ------------------------------------------------------------ 90

Db       2552 TCAGGTTCGCTTCAGATGTTAAATCTCACAAGCTTGAGTATGAATGTATTGCGTACTCCA 2611

Qy         90 ------------------------------------------------------------ 90

Db       2612 CTCGAAATTCGCAATGGTGAAAATTTGGAAGCGAGCAAACTCTATCATCGGCCAAACAGA 2671

Qy         90 ------------------------------------------------------------ 90

Db       2672 TAAGAAATTTTGGAGTTTAAAAGTTCAGTTTCCTGCAAAAATCAAACCGCGTGGAGAAAT 2731

Qy         90 ------------------------------------------------------------ 90

Db       2732 TTGTCTATGGCGGTACGTAGATATCAATAACTGCATCTGCGAAACAGAAAATTCTGGCAA 2791

Qy         90 ------------------------------------------------------------ 90

Db       2792 CCGTATCGTCCTAACTCCTAGTATCGCTGGCACATATCCATATCAGTGTGAGGGAGTTTT 2851

Qy         90 ------------------------------------------------------------ 90

Db       2852 GGGCTTGTTTTGGCTCGATGCTTCAGGAAGACAAATTACGTGGTTGTTAAGGCGGCTAAC 2911

Qy         90 ------------------------------------------------------------ 90

Db       2912 TCTACCAATCAGAAACGCTTTATTCGAAGAACCATGTTTGTTCCTCAATTCCCATCCCTA 2971

Qy         90 ------------------------------------------------------------ 90

Db       2972 CGTACAATCTGGGCTTTCCCATTGTAGTTCCTTAGGAAGTTGACTTCTTCACAACCATTC 3031

Qy         90 ------------------------------------------------------------ 90

Db       3032 TGGGATTGGTACAGTTGCAGTGGACAAGACAAATATTCATTTGCAGGCAGACTCAAATTT 3091

Qy         90 ------------------------------------------------------------ 90

Db       3092 CAATGTCTTGCCGGCTTTGAGTACTTAAATGGAGTGTTCAGGGATTGGTTTATTTGGGAC 3151

Qy         90 ------------------------------------------------------------ 90

Db       3152 TCAGGCAATGGATAAAGACAGGAATGTTTTGCAGAAAGTATTAATGTCTTTCCGGCTTTG 3211

Qy         90 ------------------------------------------------------------ 90

Db       3212 GGGACTACATACTTCTCCTACAGACAAAATCTGATGTTTAGGCAAGAGAAACTATTAACA 3271

Qy         90 ------------------------------------------------------------ 90
```

```
Db       3272 CTGATTTAATAGAGAAAGAGGAGATGGTTTCTCCTTGCGGCAATTTTATTTGTTTAGGAA 3331

Qy         90 ------------------------------------------------------------ 90

Db       3332 AGCAATTGATATGAATTGGTGTCGTAGTGTAGTTGAAATTACTAGTTAGTTTGTGTGTTT 3391

Qy         90 ------------------------------------------------------------ 90

Db       3392 AGTTTCCTTGATGTTTGATGCTTTATTCTTGGCAACCTATCTGGGTAGTATCGCCTTCTT 3451

Qy         91 --------------------AlaLysProCysLysSerLeuTrpValGlyGlyIleGly 103
                                  ||||||||||||||||||||||||||||||||||||||
Db       3452 ATCGACCTTTTCTTGTTGCAGGCAAAACCTTGTAAGAGTCTATGGGTGGGTGGAATCGGC 3511

Qy        104 ProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAsp 123
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3512 CCTAATGTCTCCAAGGATGACCTGGAGGAAGAGTTCAGCAAGTTTGGGAAAATCGAGGAT 3571

Qy        124 PheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAla 143
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3572 TTTAGGTTTCTCAGAGAACGCAAGACAGCTTTCATTGATTATTATGAGATGGATGATGCT 3631

Qy        144 LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPhe 163
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3632 TTACAGGCTAAGAGCATGAATGGAAAGCCTATGGGTGGTAGCTTTTTGCGTGTTGATTTT 3691

Qy        164 LeuArgSerGlnAlaProLysLys 171
               |||||||||||||||||||||||
Db       3692 CTCCGGTCACAAGCGCCAAAAAAA 3715
```

RESULT 6
ABA95169
ID ABA95169 standard; DNA; 349 BP.
XX
AC ABA95169;
XX
DT 20-MAY-2002 (first entry)
XX
DE Arabidopsis floral induction gene (FPA) coding fragment.
XX
KW FPA; floral induction; photoperiod; plant; flowering; FLC; FRI; frigida;
KW flowering locus C; transgenic; gene; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200212518-A2.
XX
PD 14-FEB-2002.
XX
PF 02-AUG-2001; 2001WO-US024427.
XX
PR 03-AUG-2000; 2000US-0222550P.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Amasino RM, Schomburg FM, Michaels SD, Patton D;
XX
DR WPI; 2002-227160/28.
XX
PT Novel isolated DNA sequence comprising coding sequence for floral
PT induction gene (FPA) gene which controls flowering time in plants, useful
PT for altering flowering time of plants and for downregulating flowering
PT locus C (FLC) mRNA activity.
XX
PS Claim 5; Page 37; 39pp; English.
XX
CC The invention relates to the FPA gene (floral induction promoter in
CC plants during both long and short day photoperiods) from Arabidopsis
CC thaliana. The FPA coding sequence is useful for altering flowering time
CC in a plant, and for down-regulating flowering locus C (FLC) mRNA
CC activity. The FPA gene can accelerate flowering in several late-flowering
CC mutant backgrounds and can fully compensate for addition of two naturally
CC occurring genes, frigida (FRI) and FLC which confer late-flowering
CC phenotypes. Over expression of FPA can compensate for the delaying effect
CC caused by short days on floral induction, and decreases FLC mRNA in
CC plants containing the FLC gene. Fragments of the FPA gene can also act to
CC decrease activity of an endogenous FPA gene by modifying the expression
CC of the endogenous FPA gene and expression of a portion of polypeptide
CC encoded by FPA gene can also lead to a delay of flowering in a plant. The
CC present sequence represents the A. thaliana FPA gene coding fragment
XX
SQ Sequence 349 BP; 92 A; 68 C; 81 G; 108 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.12e-16	Length:	349
Score:	466.00	Matches:	92
Percent Similarity:	94.95%	Conservative:	2
Best Local Similarity:	92.93%	Mismatches:	3
Query Match:	9.71%	Indels:	2
DB:	6	Gaps:	1

US-09-920-705-3 (1-901) x ABA95169 (1-349)

```
Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 60

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 120

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 180

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 240

Qy         81 SerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLysSerLeuTrpVal 99
              |||||||||||||||||||||||||||||||||          |||       |||::::::
Db        241 AGTCAAATTAAGATCGAATACGCACGACCGGTT------TGTTCTTATCTATATCTT 291
```

RESULT 7
ABZ21692
ID ABZ21692 standard; cDNA; 6836 BP.
XX
AC ABZ21692;
XX
DT 27-FEB-2003 (first entry)
XX
DE Human RBM15-MKL1 fusion protein encoding cDNA SEQ ID NO:1.
XX
KW Human; RBM15; RNA binding motif protein 15; megakaryoblastic leukaemia 1;
KW MKL1; fusion protein; acute megakaryoblastic leukaemia; AMKL; cytostatic;
KW t(1; 22) chromosomal rearrangement; gene therapy; chromosome 1p13;
KW chromosome 22q13; gene; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   84. .5735
FT   /*tag= a
FT   /product= "RBM15-MKL1 fusion protein"
XX
PN WO200288309-A2.
XX
PD 07-NOV-2002.
XX
PF 23-APR-2002; 2002WO-US012797.
XX
PR 27-APR-2001; 2001US-0286910P.
XX
PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PA (HOSP-) HOSPITAL FOR SICK CHILDREN.
XX

PI Morris SW, Ma Z, Hitzler JK;
XX
DR WPI; 2003-103455/09.
DR P-PSDB; ABP56065.
XX
PT New RNA-binding motif protein-15 (RBM15)-megakaryoblastic leukemia-1
PT (MKL1), MKL1-RBM15-S and MKL1-RBM15-S+AE fusion proteins, useful for
PT identifying agents useful for treating patients with acute
PT megakaryoblastic leukemia.
XX
PS Claim 2; Page 53-62; 109pp; English.
XX
CC The present invention describes an RNA-binding motif protein-15 (RBM15)-
CC megakaryoblastic leukaemia-1 (MKL1) fusion protein, a MKL1-RBM15-S fusion
CC protein, and a MKL1-RBM15-S+AE fusion protein associated with acute
CC megakaryoblastic leukaemia (AMKL). Also described: (1) an antibody that
CC specifically binds to the RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE
CC fusion proteins; (2) a non-human transgenic animal that has been altered
CC to express a gene encoding a RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE
CC fusion protein; (3) identifying an agent capable of binding to a RBM15-
CC MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE fusion protein; (4) detecting the
CC t(1;22) chromosomal rearrangement associated with AMKL; and (5) screening
CC for agents capable of (selectively) inhibiting the activity of a fusion
CC protein arising from the t(1;22) chromosomal rearrangement associated
CC with AMKL. The fusion proteins have cytostatic activity and can be used
CC in gene therapy. The RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE fusion
CC proteins and nucleotide molecules are useful for designing and preparing
CC agents that specifically inhibit the expression of the RBM15-MKL1 or MKL1
CC -RBM15 genes in cells for therapeutic and other purposes. The transgenic
CC animals are useful for identifying and testing carcinogenic or
CC therapeutic compositions. The methods are also useful for detecting the
CC t(1;22) chromosomal rearrangement associated with AMKL, or for
CC identifying agents useful for treating patients with AMKL. The antibodies
CC can be used to selectively kill cells expressing RBM15-MKL1, MKL1-RBM15-
CC S, or MKL1-RBM15-S+AE fusion protein. RBM15 is located to chromosome
CC 1p13, and MKL1 is located to chromosome 22q13. The present sequence
CC encodes the RBM15-MKL1 fusion protein of the present invention
XX
SQ Sequence 6836 BP; 1473 A; 2089 C; 1963 G; 1311 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.11e-08	Length:	6836
Score:	336.50	Matches:	237
Percent Similarity:	34.37%	Conservative:	127
Best Local Similarity:	22.38%	Mismatches:	371
Query Match:	7.01%	Indels:	324
DB:	8	Gaps:	48

US-09-920-705-3 (1-901) x ABZ21692 (1-6836)

```
Qy          7 ProPheArg-----------------AlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              |||||||||                 :::|||            :::      |||
Db       1152 CCTTTCAGAGAAGTGGATGAGATTTCACCCGAGGATGATCAGCGAGCTAACCGGACGCTC 1211

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ::::::|||:::|||           ||||||||||||||          |||   |||
Db       1212 TTCTTGGGCAACCTAGACATCACTGTAACGGAGAGTGATTTAAGAAGGGCGTTTGATCGC 1271

Qy         41 TyrGlyAspIleAspArgIleThrVal---TyrSerSerArgGly-----------Phe 55
              :::|||   |||      :::   :::         |||||||           :::
Db       1272 TTTGGAGTCATCACAGAAGTAGATATCAAGAGGCCTTCTCGCGGCCAGACTAGTACTTAC 1331

Qy         56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGly 75
                |||:::   :::   :::::::::   :::      ||||||   |||:::   |||
Db       1332 GGCTTTCTCAAATTTGAGAACTTAGATATGTCTCACCGGGCCAAATTAGCAATGTCTGGC 1391

Qy         76 AlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLys 95
                 :::      :::   |||||||||   |||   :::   |||   |||
Db       1392 AAAATTATAATTCGGAATCCTATCAAAATTGGTTATGGTAAA---GCTACACCCACCACC 1448

Qy         96 SerLeuTrpValGlyGlyIleGlyProAsnValSerLysAspAspLeuGluGluGluPhe 115
              |||||||||||||||:::||||||   |||          |||      ||||||
Db       1449 CGCCTCTGGGTGGGAGGCCTGGGACCTTGGGTTCCTCTTGCTGCCCTGGCACGAGAATTT 1508

Qy        116 SerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIle 135
                 :::||||||   |||          :::   :::          |||:::|||
Db       1509 GATCGATTTGGCACCATACGCACCATAGACTACCGAAAAGGTGATAGTTGGGCATATATC 1568

Qy        136 AspTyrTyrGluMetAspAspAlaLeuGlnAla---LysSerMetAsnGlyLysProMet 154
                |||      :::|||   |||      |||          |||   |||   |||:::
Db       1569 CAGTATGAAAGCCTGGATGCAGCGCATGCTGCCTGGACCCATATGCGGGGCTTCCCACTT 1628

Qy        155 GlyGly------SerPheLeuArgValAspPheLeuArgSerGlnAlaProLysLysGlu 172
              |||||           |||||||||||||||      ::::::          ::::::
Db       1629 GGTGGCCCAGATCGACGCCTTAGAGTAGACTTTGCCGACACCGAACATCGTTACCAGCAG 1688

Qy        173 GlnTrp-----------------------------------AlaGlySerTyrAspAsn 180
              |||:::                                   ::::::   :::
Db       1689 CAGTATCTGCAGCCTCTGCCCTTGACTCATTATGAGCTGGTGACAGATGCTTTTGGACAT 1748

Qy        181 Arg--------------AsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyr 195
              |||             |||      ::      |||      |||      |||
Db       1749 CGGGCACCAGACCCTTTGAGGGGTGCTCGGGATAGGACACCA------CCCTTACTATAC 1802

Qy        196 GluAspPheLysGlyAspValGlnProSerLysValLeuTrpIleGlyPheProProThr 215
              |||       |||:::   |||          |||:::      ||||||
Db       1803 AGAGATCGTGATAGGGACCTTTATCCTGACTCTGAT---TGGGTGCCACCCCCACCCCCA 1859

Qy        216 AlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIle 235
                        |||:::      |||
Db       1860 GTCCGA-----------GAACGCAGCACTCGGACTGCAGCTACTTCTGTGCCTGCTTAT 1907

Qy        236 GluArgValLysSerTyrProSerArg------AsnPheAlaLeuValGluPheArgSer 253
              |||   :::   |||          |||          ::::::|||          |||
Db       1908 GAGCCACTGGATAGCCTAGATCGCAGGCGGGATGGTTGGTCCTTGGACCGGGACAGAGGT 1967

Qy        254 AlaGluGluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArg 273
                    :::         ::::::                          ||||||
Db       1968 GATCGAGATCTGCCCAGCAGCAGAGAC----------------------CAGCCTAGG 2003

Qy        274 IleLysIleMetTyrSerAsnAspGluLeuProProGluGlnAsp-----------Asp 289
                 :::                    ||||||  |||                   |||
Db       2004 AAGCGA----------------AGGCTGCCTGAGGAGAGTGGAGGACGTCATCTGGAT 2045

Qy        290 ThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsnAspProSer 309
              |||      |||         |||  |||
Db       2046 AGGTCTCCTGAGAGTGACCGCCCACGAAAACGT----------------------CAC 2081

Qy        310 CysValSerSerPro---HisSerThrGlyIleProGlySerMetArgProLeuArgGly 328
              |||      ||||||   |||      ::      |||
Db       2082 TGCGCTCCTTCTCCTGACCGCAGTCCAGAATTGAGCAGTAGCCGGGATCGTTACAACAGC 2141

Qy        329 ThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValValGlyLysGluProAsn 348
                 |||:::||||||                 :::   :::::   :::   |||:::
Db       2142 GACAATGATCGATCT----------------TCCCGTCTTCTCTTGGAAAGGCCCTCT 2183

Qy        349 TrpArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIle 368
                    |||      :::|||   |||   |||
Db       2184 ---------CCAATCAGAGACGGACGAGGTAGTTTG---------------------- 2210

Qy        369 LeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrpGlu 388
                       :::||||||   :::|||      |||   |||      |||   |||
Db       2211 ------GAGAAGAGCCAGGGTGACAAGCGAGAC---CGTAAAAACTCTGCATCAGCTGAA 2261

Qy        389 GlyTyrAspProAlaGlnLeuValArgGluSerLysArgThrArgArgAspGlySerVal 408
                                                  ::::::   |||
Db       2262 --------------------------CGAGATAGGAAGCACCGGACAACTGCTCCCACT 2294

Qy        409 AspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGly---------Ser 425
              :::|||   :::|||:::      ::::::||||||   |||               |||
Db       2295 GAGGGAAAAAGCCCTCTGAAAAAAGAAGACCGCTCTGATGGGAGTGCACCTAGCACCAGC 2354

Qy        426 ValAlaAlaArgProIleArgGlyPro------------------------------- 434
              |||:::      ::::::  |||
Db       2355 ACTGCTTCCTCCAAGCTGAAGTCCCCGTCCCAGAAACAGGATGGGGGGACAGCCCCTGTG 2414
```

```
Qy       435 ------------ProAspSerAspHisIleTrpArgGlyMet------------------ 444
Db      2415 GCATCAGCCTCTCCCAAACTCTGTTTGGCCTGGCAGGGCATGCTTCTACTGAAGAACAGC 2474

Qy       445 --------------------IleAlaLysGlyGlyThrProValCysCysAlaArgCys 457
Db      2475 AACTTTCCTTCCAACATGCATCTGTTGCAGGGTGACCTCCAAGTGGCTAGTAGTCTTCTT 2534

Qy       458 Val---------ProMetGlyLysGlyIleGluThrLysLeuProGluValValAsnCys 474
Db      2535 GTGGAGGGTTCAACTGGAGGCAAAGTGGCCCAGCTCAAGATCACTCAGCGTCTCCGTTTG 2594

Qy       475 SerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAla------IleGly 492
Db      2595 GAC---CAGCCCAAGTTGGATGAAGTAACTCGACGCATCAAAGTAGCAGGGCCCAATGGT 2651

Qy       493 CysGluIleValPhePheValProAspArgGluGlu---------------------- 504
Db      2652 TATGCCATTCTTTTGGCTGTGCCTGGAAGTTCTGACAGCCGGTCCTCCTCTTCCTCAGCT 2711

Qy       505 ------AspPheAlaSerTyrThrGlu---------------PheLeuArgTyrLeuSer 517
Db      2712 GCATCAGACACTGCCACTTCTACTCAGAGGCCACTTAGGAACCTTGTGTCCTATTTAAAG 2771

Qy       518 SerLysAspArgAlaGlyValAlaLysLeuAspAspGly---------------------- 530
Db      2772 CAAAAGCAGGCAGCCGGGGTGATCAGCCTCCCTGTGGGGGGCAACAAAGACAAGGAAAAC 2831

Qy       531 ---ThrThrLeuPheLeuValProProSerAspPheLeuThrAspValLeuGln------ 547
Db      2832 ACCGGGGTCCTTCATGCCTTCCCACCTTGTGAGTTCTCCCAGCAGTTCCTGGATTCCCCT 2891

Qy       548 ---------ValThrArgGlnGluArgLeuTyrGlyValValLeuLysLeuProProPro 564
Db      2892 GCCAAGGCACTGGCCAAATCTGAAGAAGATTACCTGGTCATGATCATTGTC------CGT 2945

Qy       565 AlaValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnProLeuHisTyrMet 584
Db      2946 GCTTTGAAAAGTCCAGCCGCATTTCATGAGCAGAGAAGGAGC--------------TTG 2990

Qy       585 AspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGluAsn 604
Db      2991 GAGCGGGCCAGG-----------------------------------ACAGAGGAC 3011

Qy       605 Tyr------------IleArgGlyAlaProGluHisLeuThrAlaAlaSerLysProSer 620
Db      3012 TATCTCAAACGGAAGATTCGTTCCCGGCCGGAGAGATCGGAGCTGGTCAGGATGCACATT 3071

Qy       621 ValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSer------ 638
Db      3072 TTGGAAGAGACCTCGGCTGAGCCATCCCTCCAGGCCAAGCAGCTGAAGCTGAAGAGAGCC 3131

Qy       639 ---LeuThrProGluLeuLeuAlaThrLeuAla------------------------- 648
Db      3132 AGACTAGCCGATGACCTCAATGAGAAGATTGCACAGAGGCCTGGCCCCATGGAGCTGGTG 3191

Qy       649 ------SerIleLeuProAlaThrSerGln------------------------------ 656
Db      3192 GAGAAGAACATCCTTCCTGTTGAGTCCAGCCTGAAGGAAGCCATCATTGTGGGCCAGGTG 3251

Qy       657 ------ProAlaAlaProGluSerHisGlnProMetSerGlyProSerThrValValSer 674
Db      3252 AACTATCCCAAAGTAGCAGACAGCTCTTCCTTCGATGAGGACAGCAGCGATGCCTTATCC 3311

Qy       675 ThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArg 694
Db      3312 CCCGAGCAGCCTGCCAGCCATGAGTCCCAGGGTTCTGTGCCGTCACCCCTGGAGGCCCGA 3371

Qy       695 GlyProGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyr 714
Db      3372 GTCAGCGAACCACTGCTCAGTGCCACCTCTGCATCCCCCACCCAGGTTGTGTCTCAA--- 3428

Qy       715 ThrProAlaGlyGln-----------------------------------------LeuPro 721
```

```
Db      3429 CTTCCGATGGGCCGGGATTCCAGAGAAATGCTTTTCCTGGCAGAGCAGCCTCCTCTGCCT 3488

Qy       722 ProProProSerArgTyrProProAlaSerAsn-AsnProAsnTyrThrSerGlyMetVa 741
Db      3489 CCCCCACCTCTGCTGCCTCCCAGCCTCACCAATGGAACCA------------------- 3528

Qy       741 lHisGlyAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeuPr 761
Db      3529 ----------------------------------------------CTATCCCCACTGCC 3542

Qy       761 oAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProValSe 781
Db      3543 AAGTCCACCC-----------------------------------CCACACTCATTAAG 3566

Qy       781 r-GlnProMetValGlnGlnTyrGlnProGluAlaSerMetProAsnGlnAsnTyrGlyP 801
Db      3567 CAAAGCCA---ACCCAAGTCTGCCAGTGAGAAGTCACAGCGCAGCAAGAAGGCCAAGGAG 3623

Qy       801 ro-IleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsn 820
Db      3624 CTGAAGCCAAAGGTGAAGAAGCTCAAGTACCAC--------------------CAGTAC 3662

Qy       821 LeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGlu 840
Db      3663 ATCCCCCCGGACCAGAAGCAGGACAGGGGGGCACCCCCCATGGACTCATCCTACGCCAAG 3722

Qy       841 ---ProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThr 859
Db      3723 ATCCTGCAGCAGCAGCAGCTCTTCCTCCAG---------------------------- 3752

Qy       860 ThrAspGlyGluValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsn 879
Db      3752 ------------------------------------------------------------ 3752

Qy       880 LeuLeuLeuGlnIleGlnGlnLysGlnGlnGlnGln------------------------ 891
Db      3753 ------CTGCAGATCCTCAACCAGCAGCAGCAGCAGCACCACAACTACCAGGCCATCCTG 3806

Qy       892 ----------------------------------SerSerGlyThrPro 896
Db      3807 CCTGCCCCGCCAAAGTCAGCAGGCGAGGCCCTGGGAAGCAGCGGGACCCCC 3857
```

```
RESULT 8
AAZ51263
ID   AAZ51263 standard; cDNA; 2190 BP.
XX
AC   AAZ51263;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Human RNA-associated protein-14 (RNAAP-14) encoding cDNA.
XX
KW   RNA-associated protein; RNAAP; human; clone 1748626; cytostatic;
KW   immunosuppressive; antiinflammatory; keratolytic; neuroprotective;
KW   antiarteriosclerotic; hepatotropic; antipsoriatic; virucide; anti-HIV;
KW   antiallergic; antirheumatic; antiarthritic; opthalmological; autoimmune;
KW   antimicrobial; cell proliferative disorder; inflammation; cirrhosis;
KW   actinic keratosis; bursitis; arteriosclerosis; artherosclerosis;
KW   hepatitis; myelofibrosis; primary thrombocythemia; psoriasis; cancer;
KW   mixed connective tissue disease; MCTD; HIV; uveitis; Crohn's disease;
KW   allergy; rheumatoid arthritis; parasitic infection; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             164. .1855
FT                   /*tag=  a
FT                   /product= "Human RNA-associated protein-14"
FT                   /note= "Derived from STOMTUT02 library"
XX
PN   WO200011171-A2.
XX
```

PD 02-MAR-2000.
XX
PF 20-AUG-1999; 99WO-US019361.
XX
PR 21-AUG-1998; 98US-0097550P.
PR 12-JAN-1999; 99US-0115639P.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Hillman JL, Yue H, Tang YT, Corley NC, Guegler KJ, Gorgone GA;
PI Patterson C, Baughn MR, Lal P, Bandman O, Reddy R, Azimzai Y;
PI Shih LL, Yang J, Lu DAM;
XX
DR WPI; 2000-237651/20.
DR P-PSDB; AAY70233.
XX
PT Human RNA-associated proteins useful in diagnosing, treating and
PT preventing cell proliferative, autoimmune, inflammatory and infectious
PT disorders.
XX
PS Claim 9; Page 116-117; 123pp; English.
XX
CC The present sequence is the cDNA encoding human RNA-associated protein-14
CC (RNAAP-14), identified in Incyte clone 1748626, derived from STOMTUT02
CC library. It is expressed in reproductive, nervous, gastrointestinal, and
CC haematopoietic/immune tissues. It has cytostatic, immunosuppressive,
CC antiinflammatory, antiarteriosclerotic, hepatotropic, keratolytic,
CC neuroprotective, antipsoriatic, anti-HIV, antiallergic, antirheumatic,
CC virucide, antiarthritic, opthalmological and antimicrobial activity.
CC RNAAP antibodies are useful for diagnosis of diseases associated with
CC altered expression or activity of RNAAP. It is used to treat cell
CC proliferative, autoimmune, inflammatory and infectious disorders, like
CC actinic keratosis, bursitis, arteriosclerosis, artherosclerosis,
CC cirrhosis, hepatitis, myelofibrosis, mixed connective tissue disease
CC (MCTD), psoriasis, primary thrombocythemia and cancer, HIV, allergies,
CC rheumatoid arthritis, uveitis, Crohn's disease, and bacterial, viral and
CC parasitic infections
XX
SQ Sequence 2190 BP; 515 A; 616 C; 633 G; 426 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4.22e-07	Length:	2190
Score:	302.00	Matches:	157
Percent Similarity:	33.67%	Conservative:	78
Best Local Similarity:	22.49%	Mismatches:	232
Query Match:	6.29%	Indels:	231
DB:	3	Gaps:	27

US-09-920-705-3 (1-901) x AAZ51263 (1-2190)

```
Qy          11 AspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThr 30
               :::|||             :::   ||||||:::::::|||:::|||            :::
Db         170 GAGGATGACCAGCGGGCCACGCGCAACCTCTTCATTGGTAACCTGGACCACAGCGTATCT 229

Qy          31 GluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAspArgIleThrVal--- 49
               |||   :::|||          |||   :::||||||   |||:::   :::   :::
Db         230 GAGGTGGAGCTGCGAAGGGCCTTCGAGAAATATGGCATCATCGAGGAGGTGGTCATCAAG 289

Qy          50 TyrSerSerArgGly-----------PheAlaPheIleTyrTyrArgHisValGluGlu 65
                  :::||||||                :::|||||:::   :::::::::::::
Db         290 AGGCCTGCCCGTGGCCAGGGCGGTGCCTATGCCTTCCTCAAGTTCCAGAACCTGGACATG 349

Qy          66 AlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleLysIle 85
               |||       ||||||   |||:::   |||       :::       :::   |||||||||
Db         350 GCCCATAGGGCTAAGGTGGCCATGTCGGGCCGAGTGATTGGTCGCAACCCCATTAAGATA 409

Qy          86 GluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGlyGlyIleGlyProAsn 105
                  |||   :::   |||   |||             |||||||||||||:::|||||||||
Db         410 GGCTATGGCAAG---GCCAACCCCACCACTCGTCTCTGGGTGGGTGGCCTGGGACCTAAC 466

Qy         106 ValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArg 125
               |||       |||      ||||||   :::||||||   |||
Db         467 ACGTCACTGGCGGCTCTGGCCCGAGAGTTTGACCGCTTTGGGAGCATTCGGACCATTGAT 526

Qy         126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
               ::::::          |||:::|||   |||       :::|||   |||
Db         527 CACGTCAAAGGAGATAGCTTTGCCTATATTCAGTACGAGAGCTTGGACGCAGCCCAGGCC 586

Qy         146 Ala---LysSerMetAsnGlyLysProMetGlyGly------SerPheLeuArgValAsp 162
                  |||         |||   |||   |||:::||||||           ||||||||||||
Db         587 GCCTGTGCTAAAATGAGGGGTTTTCCCTTGGGTGGACCAGACCGCAGGCTCCGCGTGGAT 646

Qy         163 PheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsn 182
               |||   :::::::::
Db         647 TTTGCCAAAGCAGAG---------------------------------------------- 661

Qy         183 GlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
                           :::   :::||||||      |||
Db         662 --------------GAGACTCGGTACCCCCAGCAGTAC---------------------- 685

Qy         203 GlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGlu 222
               |||||||||   :::                |||                       |||
Db         686 CAGCCCTCGCCACTC--------------CCTGTGCATTATGAGCTGCTCACAGATGGA 730

Qy         223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrPro 242
                     |||             |||      ::::::   |||   ::::::   |||
Db         731 TACACCCGGCACCGCAAC------CTGGACGCCGACCTGGTGCGGGACAGGACGCCCCCA 784

Qy         243 --------------------SerArgAsnPheAlaLeuValGluPheArgSerAlaGlu 255
                                   |||   |||         ::::::   |||
Db         785 CACCTTCTGTACTCAGACCGAGACCGGACTTTTTTGGAAGGGGACTGGACCAGCCCCAGT 844

Qy         256 GluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLys 275
               ::::::          :::      |||:::|||
Db         845 AAAAGCTCTGACCGCCGAAACAGCCTTGAGGGC------------------------- 877

Qy         276 IleMetTyrSerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGly 295
                                                                   ||||||
Db         878 ----------------------------------------------------TACAGTCGC 886

Qy         296 MetLysArgSerArgThrAsp-MetPheAsnAsnAspProSerCysValSerSerProHi 315
                     |||||||||:::                  |||       |||
Db         887 TCAGTGCGCAGCCGGAGTGGTGAGCGTTGGGGGGCAGATGGAGACCGTGGTTTGCCCAAG 946

Qy         315 sSerThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAs 335
                     ||||||                              :::|||      ||
Db         947 C------------CCTGGG--------------------------AAGAGAGGCGGAAA 967

Qy         335 nGlyAlaGluTyrAsnAspValValGlyLysGluPro---------------AsnTrpAr 350
               |||       :::       ||||||   :::|||                |||
Db         968 CGGAGAAGCCTTTCCAGTGACCGTGGGAGGACAACCCATTCACCATATGAGGAACGGAGT 1027

Qy         350 gArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro----------- 366
                  |||   |||   |||:::            :::||||||   |||
Db        1028 AGGACCAAGGGCAGTGGGCAGCAGTCAGAGCGGGGCTCCGACCGCACCCCTGAGCGCAGC 1087

Qy         367 -GlyIleLeuProSerProAlaGlnGlyThrArgArg---------------------- 378
                     :::|||   :::|||   |||
Db        1088 CGCAAGGAGAACCACTCCAGTGAAGGGACCAAGGAGTCCAGCAGCAACTCCCTCAGCAAC 1147

Qy         378 ---------------------------------------------------------- 378
Db        1148 AGCAGACATGGGGCTGAGGAACGGGGCCACCACCACCACCACCACGAGGCTGCAGACTCT 1207

Qy         379 ------------------------------------------ProMetArgSerAsnPr 384
                                                        |||   |||   :::||
Db        1208 TCCCACGGGAAGAAGGCAAGAGACAGCGAGCGCAATCACCGGACCACAGAGGCCGAGCCC 1267

Qy         384 oAspSerTrpGluGlyTyrAspProAla-GlnLeuValArgGluSerLysArgThrArgA 404
               |      ||:::      :::       |||          ||||||   |||   |||
Db        1268 AAGCCTCTGGAAGAGCCAAAACACGAGACCAAAAAGCTGAAGAATCTTTCAGAGTACGCT 1327

Qy         404 rgAspGlySerValAspGlyPheThrProMetGlyValAspGluArg------------- 419
                  |||   :::                |||||:::::::::
Db        1328 CAGACACTACAGCTGGGTTGGAATGGGCTTCTGGTGTTGAAAAACAGCTGCTTCCCCACG 1387
```

```
Qy        420 -----SerPheGlyArgGlySerValAlaAla---------------------------- 428
              ::::::  ||||||
Db       1388 TCTATGCATATCCTAGAGGGGGACCAGGGGGTGATCAGCAGTCTCCTCAAAGACCACACT 1447

Qy        429 ---------------------------------ArgProIleArgGlyPro-ProAsp 436
                                                  |||     |||||| |||
Db       1448 TCTGGGAGCAAGCTGACCCAGCTGAAGATCGCCCAGCGCCTTCGACTGGACCAGCCCAAG 1507

Qy        437 SerAspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrPro--------------- 451
               |||   :::   |||   :::      |||   |||:::|||
Db       1508 CTTGACGAGGTCACACGACGCATC------AAGCAGGGGAGCCCCAACGGCTATGCGGTC 1561

Qy        452 ValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluVal 471
               :::         :::      |||   |||   |||   |||      :::|||   |||
Db       1562 CTCTTAGCCACCCAGGCAACCCCCAGTGGGCTTGGCACTGAG---GGGATGCCCACAGTA 1618

Qy        472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIle 491
                         |||       |||       |||
Db       1619 GAGCCCGGTCTGCAGAGGCGGCTTCTCAGGAACCTG------------------------ 1654

Qy        492 GlyCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGlu 511

Db       1654 ------------------------------------------------------------ 1654

Qy        512 PheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAsp--------- 528
               :::   ||||||      |||      |||||||||      |||
Db       1655 ---GTCTCCTACTTGAAACAGAAGCAGGCCGCAGGGGTGATCAGCTTGCCAGTGGGGGGG 1711

Qy        529 ------------AspGlyThr---ThrLeuPheLeuValProProSerAspPheLeuThr 543
                         |||||||||      |||:::      ||||||   ||||||
Db       1712 TCCAAGGGCAGAGACGGCACAGGCATGCTCTACGCCTTCCCACCCTGCGACTTTTCCCAG 1771

Qy        544 AspValLeuGlnValThr---------------------ArgGlnGluArgLeuTyrGly 556
                    ||||||                          :::|||   :::
Db       1772 CAGTACCTCCAGTCAGCACTAAGGACATTGGGCAAGCTAGAAGAAGAACACATGGTGATA 1831

Qy        557 ValValLeuLys---------LeuProProProAlaValProValThrAlaSerTyrArg 573
               |||:::::::::      |||   |||      |||
Db       1832 GTCATCGTCAGAGACACTGCCTAGCCCAAGCCTGTCTTTCCCAGCGTCATGTTTGTGTCA 1891

Qy        574 GlnGluSer------------------GlnSerAsnProLeuHisTyr 583
               |||:::                        |||   |||   |||:::
Db       1892 CAAAAGCAGTTATTTTAAAATCTGATCCCCTCTCTACCCTACCACTTT 1939
```

RESULT 9
ACN37956
ID ACN37956 standard; cDNA; 3127 BP.
XX
AC ACN37956;
XX
DT 18-NOV-2004 (first entry)
XX
DE Tumour-associated antigenic target (TAT) cDNA DNA324333, SEQ ID NO:1186.
XX
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2004030615-A2.
XX
PD 15-APR-2004.
XX
PF 29-SEP-2003; 2003WO-US028547.
XX
PR 02-OCT-2002; 2002US-0414971P.
XX
PA (GETH) GENENTECH INC.
XX
PI Wu TD, Zhang Z, Zhou Y;
XX
DR WPI; 2004-347921/32.
XX
PT New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
PS Claim 1; SEQ ID NO 1186; 7273pp; English.
XX
CC The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT nucleic acid of the invention
XX
SQ Sequence 3127 BP; 629 A; 1001 C; 966 G; 531 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	6.2e-07	Length:	3127
Score:	302.00	Matches:	157
Percent Similarity:	33.67%	Conservative:	78
Best Local Similarity:	22.49%	Mismatches:	232
Query Match:	6.29%	Indels:	231
DB:	13	Gaps:	27

US-09-920-705-3 (1-901) x ACN37956 (1-3127)

```
Qy         11 AspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThr 30
              :::|||            :::   ||||||::::::|||:::|||            :::
Db       1112 GAGGATGACCAGCGGGCCACGCGCAACCTCTTCATTGGTAACCTGGACCACAGCGTATCT 1171

Qy         31 GluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAspArgIleThrVal--- 49
               |||   :::|||          |||   :::||||||   |||:::   :::   :::
Db       1172 GAGGTGGAGCTGCGAAGGGCCTTCGAGAAATATGGCATCATCGAGGAGGTGGTCATCAAG 1231

Qy         50 TyrSerSerArgGly------------PheAlaPheIleTyrTyrArgHisValGluGlu 65
                    :::||||||              :::||||||:::   ::::::::::::::
Db       1232 AGGCCTGCCCGTGGCCAGGGCGGTGCCTATGCCTTCCTCAAGTTCCAGAACCTGGACATG 1291

Qy         66 AlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleLysIle 85
               |||      ||||||   |||:::   |||      :::      :::   |||||||||
Db       1292 GCCCATAGGGCTAAGGTGGCCATGTCGGGCCGAGTGATTGGTCGCAACCCCATTAAGATA 1351

Qy         86 GluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGlyGlyIleGlyProAsn 105
                  :::   |||   |||          |||||||||||||:::|||||||||
Db       1352 GGCTATGGCAAG---GCCAACCCCACCACTCGTCTCTGGGTGGGTGGCCTGGGACCTAAC 1408

Qy        106 ValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArg 125
                  |||      |||      ||||||   :::||||||   |||
Db       1409 ACGTCACTGGCGGCTCTGGCCCGAGAGTTTGACCGCTTTGGGAGCATTCGGACCATTGAT 1468

Qy        126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
                  ::::::            |||:::|||   |||      :::|||   |||
Db       1469 CACGTCAAAGGAGATAGCTTTGCCTATATTCAGTACGAGAGCTTGGACGCAGCCCAGGCC 1528
```

```
Qy       146 Ala---LysSerMetAsnGlyLysProMetGlyGly------SerPheLeuArgValAsp 162
Db      1529 GCCTGTGCTAAAATGAGGGGTTTTCCCTTGGGTGGACCAGACCGCAGGCTCCGCGTGGAT 1588

Qy       163 PheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsn 182
Db      1589 TTTGCCAAAGCAGAG---------------------------------------------- 1603

Qy       183 GlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
Db      1604 --------------GAGACTCGGTACCCCCAGCAGTAC---------------------- 1627

Qy       203 GlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGlu 222
Db      1628 CAGCCCTCGCCACTC--------------CCTGTGCATTATGAGCTGCTCACAGATGGA 1672

Qy       223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrPro 242
Db      1673 TACACCCGGCACCGCAAC------CTGGACGCCGACCTGGTGCGGGACAGGACGCCCCCA 1726

Qy       243 --------------------SerArgAsnPheAlaLeuValGluPheArgSerAlaGlu 255
Db      1727 CACCTTCTGTACTCAGACCGAGACCGGACTTTTTTGGAAGGGGACTGGACCAGCCCCAGT 1786

Qy       256 GluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLys 275
Db      1787 AAAAGCTCTGACCGCCGAAACAGCCTTGAGGGC------------------------- 1819

Qy       276 IleMetTyrSerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGly 295
Db      1820 ----------------------------------------------------TACAGTCGC 1828

Qy       296 MetLysArgSerArgThrAsp-MetPheAsnAsnAspProSerCysValSerSerProHi 315
Db      1829 TCAGTGCGCAGCCGGAGTGGTGAGCGTTGGGGGGCAGATGGAGACCGTGGTTTGCCCAAG 1888

Qy       315 sSerThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAs 335
Db      1889 C------------CCTGGG-------------------------AAGAGAGGCGGAAA 1909

Qy       335 nGlyAlaGluTyrAsnAspValValGlyLysGluPro---------------AsnTrpAr 350
Db      1910 CGGAGAAGCCTTTCCAGTGACCGTGGGAGGACAACCCATTCACCATATGAGGAACGGAGT 1969

Qy       350 gArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro----------- 366
Db      1970 AGGACCAAGGGCAGTGGGCAGCAGTCAGAGCGGGGCTCCGACCGCACCCCTGAGCGCAGC 2029

Qy       367 -GlyIleLeuProSerProAlaGlnGlyThrArgArg---------------------- 378
Db      2030 CGCAAGGAGAACCACTCCAGTGAAGGGACCAAGGAGTCCAGCAGCAACTCCCTCAGCAAC 2089

Qy       378 ------------------------------------------------------------ 378
Db      2090 AGCAGACATGGGGCTGAGGAACGGGGCCACCACCACCACCACCACGAGGCTGCAGACTCT 2149

Qy       379 ----------------------------------------ProMetArgSerAsnPr 384
Db      2150 TCCCACGGGAAGAAGGCAAGAGACAGCGAGCGCAATCACCGGACCACAGAGGCCGAGCCC 2209

Qy       384 oAspSerTrpGluGlyTyrAspProAla-GlnLeuValArgGluSerLysArgThrArgA 404
Db      2210 AAGCCTCTGGAAGAGCCAAAACACGAGACCAAAAAGCTGAAGAATCTTTCAGAGTACGCT 2269

Qy       404 rgAspGlySerValAspGlyPheThrProMetGlyValAspGluArg------------- 419
Db      2270 CAGACACTACAGCTGGGTTGGAATGGGCTTCTGGTGTTGAAAAACAGCTGCTTCCCCACG 2329

Qy       420 -----SerPheGlyArgGlySerValAlaAla--------------------------- 428
Db      2330 TCTATGCATATCCTAGAGGGGGACCAGGGGGTGATCAGCAGTCTCCTCAAAGACCACACT 2389

Qy       429 ----------------------------------ArgProIleArgGlyPro-ProAsp 436
```

```
Db      2390 TCTGGGAGCAAGCTGACCCAGCTGAAGATCGCCCAGCGCCTTCGACTGGACCAGCCCAAG 2449

Qy       437 SerAspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrPro--------------- 451
Db      2450 CTTGACGAGGTCACACGACGCATC------AAGCAGGGGAGCCCCAACGGCTATGCGGTC 2503

Qy       452 ValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluVal 471
Db      2504 CTCTTAGCCACCCAGGCAACCCCCAGTGGGCTTGGCACTGAG---GGGATGCCCACAGTA 2560

Qy       472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIle 491
Db      2561 GAGCCCGGTCTGCAGAGGCGGCTTCTCAGGAACCTG---------------------- 2596

Qy       492 GlyCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGlu 511
Db      2596 ------------------------------------------------------------ 2596

Qy       512 PheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAsp--------- 528
Db      2597 ---GTCTCCTACTTGAAACAGAAGCAGGCCGCAGGGGTGATCAGCTTGCCAGTGGGGGGG 2653

Qy       529 -----------AspGlyThr---ThrLeuPheLeuValProProSerAspPheLeuThr 543
Db      2654 TCCAAGGGCAGAGACGGCACAGGCATGCTCTACGCCTTCCCACCCTGCGACTTTTCCCAG 2713

Qy       544 AspValLeuGlnValThr---------------------ArgGlnGluArgLeuTyrGly 556
Db      2714 CAGTACCTCCAGTCAGCACTAAGGACATTGGGCAAGCTAGAAGAAGAACACATGGTGATA 2773

Qy       557 ValValLeuLys---------LeuProProProAlaValProValThrAlaSerTyrArg 573
Db      2774 GTCATCGTCAGAGACACTGCCTAGCCCAAGCCTGTCTTTCCCAGCGTCATGTTTGTGTCA 2833

Qy       574 GlnGluSer------------------GlnSerAsnProLeuHisTyr 583
Db      2834 CAAAAGCAGTTATTTTAAAATCTGATCCCCTCTCTACCCTACCACTTT 2881
```

RESULT 10
AAD53223_4
Continuation (5 of 5) of AAD53223 from base 400001 (Human chromosome 3 p-arm breakpoint

WP Sequence split into 5 fragments LOCUS AAD53223 Accession Aad53223

	Fragment Name	Begin	End
WP	AAD53223_0	1	110000
WP	AAD53223_1	100001	210000
WP	AAD53223_2	200001	310000
WP	AAD53223_3	300001	410000
WP	AAD53223_4	400001	487980

Alignment Scores:

Pred. No.:	2.26e-05	Length:	87980
Score:	302.00	Matches:	157
Percent Similarity:	33.67%	Conservative:	78
Best Local Similarity:	22.49%	Mismatches:	232
Query Match:	6.29%	Indels:	231
DB:	8	Gaps:	27

US-09-920-705-3 (1-901) x AAD53223_4 (1-87980)

```
Qy        11 AspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThr 30
Db     73903 GAGGATGACCAGCGGGCCACGCGCAACCTCTTCATTGGTAACCTGGACCACAGCGTATCT 73962

Qy        31 GluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAspArgIleThrVal--- 49
Db     73963 GAGGTGGAGCTGCGAAGGGCCTTCGAGAAATATGGCATCATCGAGGAGGTGGTCATCAAG 74022

Qy        50 TyrSerSerArgGly------------PheAlaPheIleTyrTyrArgHisValGluGlu 65
Db     74023 AGGCCTGCCCGTGGCCAGGGCGGTGCCTATGCCTTCCTCAAGTTCCAGAACCTGGACATG 74082

Qy        66 AlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleLysIle 85
```

```
Db      74083 GCCCATAGGGCTAAGGTGGCCATGTCGGGCCGAGTGATTGGTCGCAACCCCATTAAGATA 74142

Qy        86 GluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGlyGlyIleGlyProAsn 105

Db      74143 GGCTATGGCAAG---GCCAACCCCACCACTCGTCTCTGGGTGGGTGGCCTGGGACCTAAC 74199

Qy       106 ValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArg 125

Db      74200 ACGTCACTGGCGGCTCTGGCCCGAGAGTTTGACCGCTTTGGGAGCATTCGGACCATTGAT 74259

Qy       126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145

Db      74260 CACGTCAAAGGAGATAGCTTTGCCTATATTCAGTACGAGAGCTTGGACGCAGCCCAGGCC 74319

Qy       146 Ala---LysSerMetAsnGlyLysProMetGlyGly------SerPheLeuArgValAsp 162

Db      74320 GCCTGTGCTAAAATGAGGGGTTTTCCCTTGGGTGGACCAGACCGCAGGCTCCGCGTGGAT 74379

Qy       163 PheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsn 182

Db      74380 TTTGCCAAAGCAGAG---------------------------------------------- 74394

Qy       183 GlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202

Db      74395 --------------GAGACTCGGTACCCCCAGCAGTAC--------------------- 74418

Qy       203 GlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGlu 222

Db      74419 CAGCCCTCGCCACTC--------------CCTGTGCATTATGAGCTGCTCACAGATGGA 74463

Qy       223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrPro 242

Db      74464 TACACCCGGCACCGCAAC------CTGGACGCCGACCTGGTGCGGGACAGGACGCCCCCA 74517

Qy       243 --------------------SerArgAsnPheAlaLeuValGluPheArgSerAlaGlu 255

Db      74518 CACCTTCTGTACTCAGACCGAGACCGGACTTTTTTGGAAGGGGACTGGACCAGCCCCAGT 74577

Qy       256 GluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLys 275

Db      74578 AAAAGCTCTGACCGCCGAAACAGCCTTGAGGGC---------------------------- 74610

Qy       276 IleMetTyrSerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGly 295

Db      74611 ------------------------------------------------------TACAGTCGC 74619

Qy       296 MetLysArgSerArgThrAsp-MetPheAsnAsnAspProSerCysValSerSerProHi 315

Db      74620 TCAGTGCGCAGCCGGAGTGGTGAGCGTTGGGGGGCAGATGGAGACCGTGGTTTGCCCAAG 74679

Qy       315 sSerThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAs 335

Db      74680 C------------CCTGGG--------------------------AAGAGAGGCGGAAA 74700

Qy       335 nGlyAlaGluTyrAsnAspValValGlyLysGluPro---------------AsnTrpAr 350

Db      74701 CGGAGAAGCCTTTCCAGTGACCGTGGGAGGACAACCCATTCACCATATGAGGAACGGAGT 74760

Qy       350 gArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro---------- 366

Db      74761 AGGACCAAGGGCAGTGGGCAGCAGTCAGAGCGGGGCTCCGACCGCACCCCTGAGCGCAGC 74820

Qy       367 -GlyIleLeuProSerProAlaGlnGlyThrArgArg---------------------- 378

Db      74821 CGCAAGGAGAACCACTCCAGTGAAGGGACCAAGGAGTCCAGCAGCAACTCCCTCAGCAAC 74880

Qy       378 ------------------------------------------------------------ 378

Db      74881 AGCAGACATGGGGCTGAGGAACGGGGCCACCACCACCACCACGAGGCTGCAGACTCT 74940

Qy       379 ----------------------------------------ProMetArgSerAsnPr 384
```

```
Db      74941 TCCCACGGGAAGAAGGCAAGAGACAGCGAGCGCAATCACCGGACCACAGAGGCCGAGCCC 75000

Qy         384 oAspSerTrpGluGlyTyrAspProAla-GlnLeuValArgGluSerLysArgThrArgA 404

Db      75001 AAGCCTCTGGAAGAGCCAAAACACGAGACCAAAAAGCTGAAGAATCTTTCAGAGTACGCT 75060

Qy         404 rgAspGlySerValAspGlyPheThrProMetGlyValAspGluArg------------ 419

Db      75061 CAGACACTACAGCTGGGTTGGAATGGGCTTCTGGTGTTGAAAAACAGCTGCTTCCCCACG 75120

Qy         420 -----SerPheGlyArgGlySerValAlaAla---------------------------- 428

Db      75121 TCTATGCATATCCTAGAGGGGGACCAGGGGGTGATCAGCAGTCTCCTCAAAGACCACACT 75180

Qy         429 ----------------------------------ArgProIleArgGlyPro-ProAsp 436

Db      75181 TCTGGGAGCAAGCTGACCCAGCTGAAGATCGCCCAGCGCCTTCGACTGGACCAGCCCAAG 75240

Qy         437 SerAspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrPro--------------- 451

Db      75241 CTTGACGAGGTCACACGACGCATC------AAGCAGGGGAGCCCCAACGGCTATGCGGTC 75294

Qy         452 ValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluVal 471

Db      75295 CTCTTAGCCACCCAGGCAACCCCCAGTGGGCTTGGCACTGAG---GGGATGCCCACAGTA 75351

Qy         472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIle 491

Db      75352 GAGCCCGGTCTGCAGAGGCGGCTTCTCAGGAACCTG---------------------- 75387

Qy         492 GlyCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGlu 511

Db      75387 ------------------------------------------------------------ 75387

Qy         512 PheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAsp--------- 528

Db      75388 ---GTCTCCTACTTGAAACAGAAGCAGGCCGCAGGGGTGATCAGCTTGCCAGTGGGGGGG 75444

Qy         529 -----------AspGlyThr---ThrLeuPheLeuValProProSerAspPheLeuThr 543

Db      75445 TCCAAGGGCAGAGACGGCACAGGCATGCTCTACGCCTTCCCACCCTGCGACTTTTCCCAG 75504

Qy         544 AspValLeuGlnValThr---------------------ArgGlnGluArgLeuTyrGly 556

Db      75505 CAGTACCTCCAGTCAGCACTAAGGACATTGGGCAAGCTAGAAGAAGAACACATGGTGATA 75564

Qy         557 ValValLeuLys---------LeuProProProAlaValProValThrAlaSerTyrArg 573

Db      75565 GTCATCGTCAGAGACACTGCCTAGCCCAAGCCTGTCTTTCCCAGCGTCATGTTTGTGTCA 75624

Qy         574 GlnGluSer-----------------GlnSerAsnProLeuHisTyr 583

Db      75625 CAAAAGCAGTTATTTTAAAATCTGATCCCCTCTCTACCCTACCACTTT 75672
```

```
RESULT 11
ADS95035
ID   ADS95035 standard; DNA; 2673 BP.
XX
AC   ADS95035;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Human therapeutic DNA sequence #74, SEQ ID 74.
XX
KW   Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;
KW   Gene therapy; proliferative disorder; cancer; inflammatory disorder;
KW   immune disorder; bacterial disorder; viral disorder; human; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004039319-A2.
XX
PD   13-MAY-2004.
```

XX
PF 24-OCT-2003; 2003WO-US033657.
XX
PR 25-OCT-2002; 2002US-0421080P.
PR 25-OCT-2002; 2002US-0421614P.
PR 30-OCT-2002; 2002US-0422178P.
PR 15-NOV-2002; 2002US-0426355P.
PR 15-NOV-2002; 2002US-0426384P.
PR 27-NOV-2002; 2002US-0429224P.
PR 27-NOV-2002; 2002US-0429275P.
PR 04-DEC-2002; 2002US-0430657P.
PR 04-DEC-2002; 2002US-0430663P.
PR 04-DEC-2002; 2002US-0430684P.
PR 05-DEC-2002; 2002US-0430965P.
PR 05-DEC-2002; 2002US-0431458P.
PR 12-DEC-2002; 2002US-0433500P.
PR 13-DEC-2002; 2002US-0433318P.
PR 03-JAN-2003; 2003US-0437914P.
PR 17-JAN-2003; 2003US-0440820P.
PR 18-APR-2003; 2003US-0463700P.
PR 18-APR-2003; 2003US-0463732P.
PR 02-MAY-2003; 2003US-0467203P.
PR 02-MAY-2003; 2003US-0467230P.
PR 19-MAY-2003; 2003US-0471306P.
PR 22-MAY-2003; 2003US-0472420P.
PR 09-JUN-2003; 2003US-0476609P.
PR 09-JUN-2003; 2003US-0476632P.
PR 08-JUL-2003; 2003US-0485223P.
PR 08-JUL-2003; 2003US-0485325P.
PR 08-JUL-2003; 2003US-0485359P.
PR 14-JUL-2003; 2003US-0486480P.
PR 15-JUL-2003; 2003US-0486960P.
PR 08-AUG-2003; 2003US-0493370P.
PR 08-AUG-2003; 2003US-0493573P.
XX
PA (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI Williams LT, Chu K, Lee E, Hestir K;
XX
DR WPI; 2004-376048/35.
DR P-PSDB; ADS95116.
XX
PT New nucleic acid molecules, useful for treating proliferative disorders,
PT e.g. cancer, and inflammatory, immune, bacterial or viral disorders.
XX
PS Disclosure; SEQ ID NO 74; 220pp; English.
XX
CC The present invention relates to novel human proteins (ADS95043-ADS95123)
CC and DNA sequences (ADS94962-ADS95042 and ADS95124-ADS95204). Also
CC disclosed are murine DNA sequences (ADS95205-ADS95285). The sequences are
CC useful for treating proliferative disorders, e.g. cancer, and
CC inflammatory, immune, bacterial or viral disorders. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences. Also, Claim 1 of this
CC specification refers to SEQ IDs 419-627 and Claim 14 refers to SEQ ID 210
CC -418, however, the sequence listing comprises 324 sequences.
XX
SQ Sequence 2673 BP; 540 A; 883 C; 858 G; 392 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	9.65e-07	Length:	2673
Score:	297.00	Matches:	148
Percent Similarity:	33.90%	Conservative:	71
Best Local Similarity:	22.91%	Mismatches:	212
Query Match:	6.19%	Indels:	215
DB:	13	Gaps:	24

US-09-920-705-3 (1-901) x ADS95035 (1-2673)

```
Qy         11 AspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThr 30
              :::|||          :::   ||||||::::::|||:::|||           :::
Db        988 GAGGATGACCAGCGGGCCACGCGCAACCTCTTCATTGGTAACCTGGACCACAGCGTATCT 1047

Qy         31 GluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAspArgIleThrVal--- 49
              |||   :::|||          |||   :::||||||    |||:::   :::   :::
Db       1048 GAGGTGGAGCTGCGAAGGGCCTTCGAGAAATATGGCATCATCGAGGAGGTGGTCATCAAG 1107

Qy         50 TyrSerSerArgGly-----------PheAlaPheIleTyrTyrArgHisValGluGlu 65
                 :::||||||              :::||||||:::   ::::::::::::::
Db       1108 AGGCCTGCCCGTGGCCAGGGCGGTGCCTATGCCTTCCTCAAGTTCCAGAACCTGGACATG 1167

Qy         66 AlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleLysIle 85
              |||      ||||||   |||:::   |||       :::      :::   |||||||||
Db       1168 GCCCATAGGGCTAAGGTGGCCATGTCGGGCCGAGTGATTGGTCGCAACCCCATTAAGATA 1227

Qy         86 GluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGlyGlyIleGlyProAsn 105
                 |||   :::   |||   |||          ||||||||||||||:::|||||||||
Db       1228 GGCTATGGCAAG---GCCAACCCCACCACTCGTCTCTGGGTGGGTGGCCTGGGACCTAAC 1284

Qy        106 ValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArg 125
              |||         |||      ||||||   :::||||||   |||
Db       1285 ACGTCACTGGCGGCTCTGGCCCGAGAGTTTGACCGCTTTGGGAGCATTCGGACCATTGAT 1344

Qy        126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
                 ::::::           |||:::|||   |||      :::|||   |||
Db       1345 CACGTCAAAGGAGATAGCTTTGCCTATATTCAGTACGAGAGCTTGGACGCAGCCCAGGCC 1404

Qy        146 Ala---LysSerMetAsnGlyLysProMetGlyGly------SerPheLeuArgValAsp 162
              |||       |||   |||   |||:::||||||             ||||||||||||
Db       1405 GCCTGTGCTAAAATGAGGGGTTTTCCCTTGGGTGGACCAGACCGCAGGCTCCGCGTGGAT 1464

Qy        163 PheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsn 182
              |||   :::::::::
Db       1465 TTTGCCAAAGCAGAG---------------------------------------------- 1479

Qy        183 GlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
                           :::   :::|||||||      |||
Db       1480 --------------GAGACTCGGTACCCCCAGCAGTAC--------------------- 1503

Qy        203 GlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGlu 222
              |||||||||   :::               |||                  |||
Db       1504 CAGCCCTCGCCACTC--------------CCTGTGCATTATGAGCTGCTCACAGATGGA 1548

Qy        223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrPro 242
                    |||             |||       ::::::   |||   ::::::   |||
Db       1549 TACACCCGGCACCGCAAC------CTGGACGCCGACCTGGTGCGGGACAGGACGCCCCCA 1602

Qy        243 --------------------SerArgAsnPheAlaLeuValGluPheArgSerAlaGlu 255
                                  |||   |||         ::::::   |||
Db       1603 CACCTTCTGTACTCAGACCGAGACCGGACTTTTTTGGAAGGGGACTGGACCAGCCCCAGT 1662

Qy        256 GluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLys 275
              ::::::         :::      |||:::|||
Db       1663 AAAAGCTCTGACCGCCGAAACAGCCTTGAGGGC-------------------------- 1695

Qy        276 IleMetTyrSerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGly 295
                                                                  ||||||
Db       1696 ----------------------------------------------------TACAGTCGC 1704

Qy        296 MetLysArgSerArgThrAsp-MetPheAsnAsnAspProSerCysValSerSerProHi 315
                    ||||||||:::                          |||      |||
Db       1705 TCAGTGCGCAGCCGGAGTGGTGAGCGTTGGGGGGCAGATGGAGACCGTGGTTTGCCCAAG 1764

Qy        315 sSerThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAs 335
                          ||||||                        :::|||      ||
Db       1765 C------------CCTGGG-------------------------AAGAGAGGCGGAAA 1785

Qy        335 nGlyAlaGluTyrAsnAspValValGlyLysGluPro--------------AsnTrpAr 350
              ||||       :::         ||||||   :::|||               |||
Db       1786 CGGAGAAGCCTTTCCAGTGACCGTGGGAGGACAACCCATTCACCATATGAGGAACGGAGT 1845

Qy        350 gArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro---------- 366
                 |||   |||   |||:::          :::||||||   |||
Db       1846 AGGACCAAGGGCAGTGGGCAGCAGTCAGAGCGGGGCTCCGACCGCACCCCTGAGCGCAGC 1905
```

```
Qy        367 -GlyIleLeuProSerProAlaGlnGlyThrArgArg---------------------- 378
Db       1906 CGCAAGGAGAACCACTCCAGTGAAGGGACCAAGGAGTCCAGCAGCAACTCCCTCAGCAAC 1965

Qy        378 ------------------------------------------------------------ 378
Db       1966 AGCAGACATGGGGCTGAGGAACGGGGCCACCACCACCACCACCACGAGGCTGCAGACTCT 2025

Qy        379 ----------------------------------------ProMetArgSerAsnPr 384
Db       2026 TCCCACGGGAAGAAGGCAAGAGACAGCGAGCGCAATCACCGGACCACAGAGGCCGAGCCC 2085

Qy        384 oAspSerTrpGluGlyTyrAspProAla-GlnLeuValArgGluSerLysArgThrArgA 404
Db       2086 AAGCCTCTGGAAGAGCCAAAACACGAGACCAAAAAGCTGAAGAATCTTTCAGAGTACGCT 2145

Qy        404 rgAspGlySerValAspGlyPheThrProMetGlyValAspGluArg------------- 419
Db       2146 CAGACACTACAGCTGGGTTGGAATGGGCTTCTGGTGTTGAAAAACAGCTGCTTCCCCACG 2205

Qy        420 -----SerPheGlyArgGlySerValAlaAla---------------------------- 428
Db       2206 TCTATGCATATCCTAGAGGGGGACCAGGGGGTGATCAGCAGTCTCCTCAAAGACCACACT 2265

Qy        429 ----------------------------------ArgProIleArgGlyPro-ProAsp 436
Db       2266 TCTGGGAGCAAGCTGACCCAGCTGAAGATCGCCCAGCGCCTTCGACTGGACCAGCCCAAG 2325

Qy        437 SerAspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrPro--------------- 451
Db       2326 CTTGACGAGGTCACACGACGCATC------AAGCAGGGGAGCCCCAACGGCTATGCGGTC 2379

Qy        452 ValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluVal 471
Db       2380 CTCTTAGCCACCCAGGCAACCCCCAGTGGGCTTGGCACTGAG---GGGATGCCCACAGTA 2436

Qy        472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIle 491
Db       2437 GAGCCCGGTCTGCAGAGGCGGCTTCTCAGGAACCTG---------------------- 2472

Qy        492 GlyCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGlu 511
Db       2472 ------------------------------------------------------------ 2472

Qy        512 PheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAsp--------- 528
Db       2473 ---GTCTCCTACTTGAAACAGAAGCAGGCCGCAGGGGTGATCAGCTTGCCAGTGGGGGGG 2529

Qy        529 ------------AspGlyThr---ThrLeuPheLeuValProProSerAspPheLeuThr 543
Db       2530 TCCAAGGGCAGAGACGGCACAGGCATGCTCTACGCCTTCCCACCCTGCGACTTTTCCCAG 2589

Qy        544 AspValLeuGln 547
Db       2590 CAGTACCTCCAG 2601
```

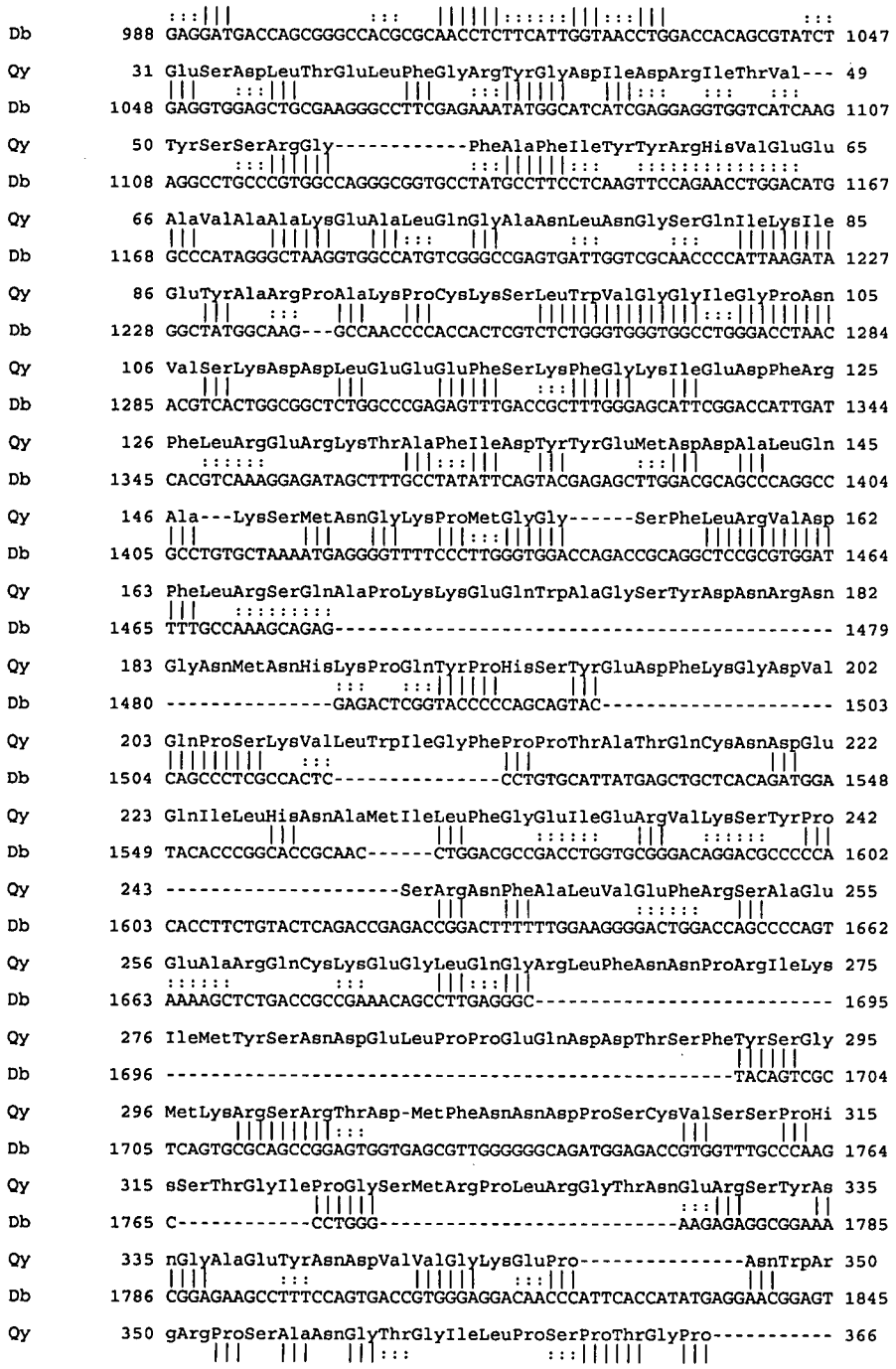
RESULT 12
ADS95197
ID ADS95197 standard; DNA; 2673 BP.
XX
AC ADS95197;
XX
DT 02-DEC-2004 (first entry)
XX
DE Human therapeutic DNA sequence #74, SEQ ID 236.
XX
KW Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;
KW Gene therapy; proliferative disorder; cancer; inflammatory disorder;
KW immune disorder; bacterial disorder; viral disorder; human; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2004039319-A2.
XX
PD 13-MAY-2004.
XX
PF 24-OCT-2003; 2003WO-US033657.
XX
PR 25-OCT-2002; 2002US-0421080P.
PR 25-OCT-2002; 2002US-0421614P.
PR 30-OCT-2002; 2002US-0422178P.
PR 15-NOV-2002; 2002US-0426355P.
PR 15-NOV-2002; 2002US-0426384P.
PR 27-NOV-2002; 2002US-0429224P.
PR 27-NOV-2002; 2002US-0429275P.
PR 04-DEC-2002; 2002US-0430657P.
PR 04-DEC-2002; 2002US-0430663P.
PR 04-DEC-2002; 2002US-0430684P.
PR 05-DEC-2002; 2002US-0430965P.
PR 05-DEC-2002; 2002US-0431458P.
PR 12-DEC-2002; 2002US-0433500P.
PR 13-DEC-2002; 2002US-0433318P.
PR 03-JAN-2003; 2003US-0437914P.
PR 17-JAN-2003; 2003US-0440820P.
PR 18-APR-2003; 2003US-0463700P.
PR 18-APR-2003; 2003US-0463732P.
PR 02-MAY-2003; 2003US-0467203P.
PR 02-MAY-2003; 2003US-0467230P.
PR 19-MAY-2003; 2003US-0471306P.
PR 22-MAY-2003; 2003US-0472420P.
PR 09-JUN-2003; 2003US-0476609P.
PR 09-JUN-2003; 2003US-0476632P.
PR 08-JUL-2003; 2003US-0485223P.
PR 08-JUL-2003; 2003US-0485325P.
PR 08-JUL-2003; 2003US-0485359P.
PR 14-JUL-2003; 2003US-0486480P.
PR 15-JUL-2003; 2003US-0486960P.
PR 08-AUG-2003; 2003US-0493370P.
PR 08-AUG-2003; 2003US-0493573P.
XX
PA (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI Williams LT, Chu K, Lee E, Hestir K;
XX
DR WPI; 2004-376048/35.
DR P-PSDB; ADS95116.
XX
PT New nucleic acid molecules, useful for treating proliferative disorders,
PT e.g. cancer, and inflammatory, immune, bacterial or viral disorders.
XX
PS Disclosure; SEQ ID NO 236; 220pp; English.
XX
CC The present invention relates to novel human proteins (ADS95043-ADS95123)
CC and DNA sequences (ADS94962-ADS95042 and ADS95124-ADS95204). Also
CC disclosed are murine DNA sequences (ADS95205-ADS95285). The sequences are
CC useful for treating proliferative disorders, e.g. cancer, and
CC inflammatory, immune, bacterial or viral disorders. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences. Also, Claim 1 of this
CC specification refers to SEQ IDs 419-627 and Claim 14 refers to SEQ ID 210
CC -418, however, the sequence listing comprises 324 sequences.
XX
SQ Sequence 2673 BP; 540 A; 883 C; 858 G; 392 T; 0 U; 0 Other;

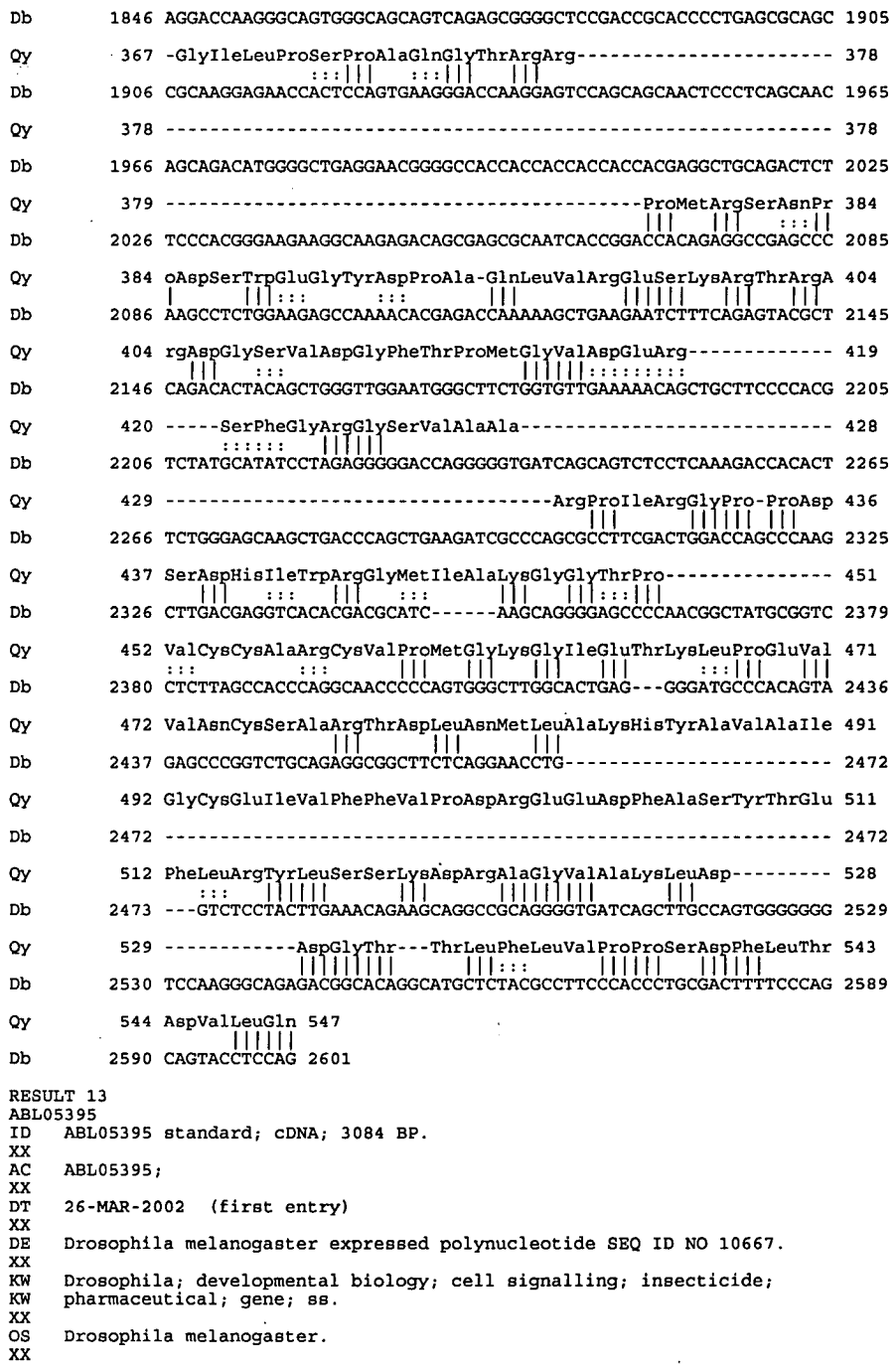
Alignment Scores:

Pred. No.:	9.65e-07	Length:	2673
Score:	297.00	Matches:	148
Percent Similarity:	33.90%	Conservative:	71
Best Local Similarity:	22.91%	Mismatches:	212
Query Match:	6.19%	Indels:	215
DB:	13	Gaps:	24

US-09-920-705-3 (1-901) x ADS95197 (1-2673)

```
Qy         11 AspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThr 30
```





RESULT 13
ABL05395
ID ABL05395 standard; cDNA; 3084 BP.
XX
AC ABL05395;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10667.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX

PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB61292.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 10667; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3084 BP; 706 A; 868 C; 885 G; 625 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3.6e-06	Length:	3084
Score:	287.50	Matches:	140
Percent Similarity:	32.70%	Conservative:	65
Best Local Similarity:	22.33%	Mismatches:	201
Query Match:	5.99%	Indels:	221
DB:	4	Gaps:	27

US-09-920-705-3 (1-901) x ABL05395 (1-3084)

```
Qy          11 AspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThr 30
               :::|||           :::    |||:::   |||:::|||
Db        1111 GAGGATGATCCTCTTTCTACGCGCACACTGTTTGCAGGAAATCTGGAGGTTACCATCGCC 1170

Qy          31 GluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAspArgIleThrValTyr 50
               :::   :::|||      :::||||||:::|||||||   :::|||   |||   :::
Db        1171 GACGACGAACTGCGTCGCATCTTTGGCAAATACGGAGTGGTAGATGACATCGACATCAAG 1230

Qy          51 ---------------SerSerArgGlyPheAlaPheIleTyrTyrArgHisValGluGlu 65
                            :::         |||||||||:::   |||:::::::::::::
Db        1231 CGACCGCCGCCGGGTACGGGGAACGCGTTCGCGTTTGTACGTTACCAAAACCTGGATATG 1290

Qy          66 AlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleLysIle 85
               |||      ||||||      |||   |||       :::       |||   ||||||
Db        1291 GCTCACCGAGCGAAGATCGAATTGTCTGGACAGTACATTGGCAAGTTTCAGTGCAAGATC 1350

Qy          86 GluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGlyGlyIleGlyProAsn 105
               |||   :::          |||          :::|||:::||||||:::|||
Db        1351 GGATACGGGAAA---GTGACTCCGGCAACTCGTATGTGGATCGGAGGCTTAGGTGCCTGG 1407

Qy         106 ValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArg 125
               |||          ||||||   ||||||   :::||||||   |||:::
Db        1408 ACTTCTGTCACTCAGCTGGAGCGAGAGTTTGATCGCTTCGGTGCGATCAAGAAGATTGAA 1467

Qy         126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
               :::   :::           |||:::|||   |||       ::::::   |||
Db        1468 TACCAGAAGGGCGAGCCGTACGCTTATATTCAGTACGAGACAGTGGAGGCAGCCACGGCG 1527

Qy         146 Ala---LysSerMetAsnGlyLysProMetGlyGly------SerPheLeuArgValAsp 162
               |||   |||   |||   |||   |||:::||||||            ||||||   |||
Db        1528 GCGGTAAAGGAGATGCGAGGATTCCCATTGGGTGGACCGGAGCGTCGACTCCGCACGGAC 1587

Qy         163 Phe--------------------------------LeuArgSerGlnAlaProLysLys 171
               |||                                     :::|||       |||
Db        1588 TTTGCCGAGTTGCCCGGCGCCACACCGGCAGCGCCTTTCAAGAGCTCAAAGCCGCCATAT 1647

Qy         172 GluGlnTrpAlaGlySerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGlnTyr 191
               ::::::   |||       |||                                 :::|||:::|||
Db        1648 GATGAAAGTGCGTTGGAGTAC----------------------AGACGTCCGGAATAC 1683

Qy         192 ProHisSerTyrGluAspPhe---------------LysGlyAspValGlnProSerLys 206
                     ||||||:::                        :::|||       |||
Db        1684 GATCCGTATTACGAAGAGTCCGCCGCCTACGCTCCGCGCGGTGGCTATTCTCCG------ 1737

Qy         207 ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHis 226
                        :::|||||||
Db        1738 --------------TATCCACCTAGGGGT---------------------------- 1752

Qy         227 AsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsnPhe 246
                                                        |||       |||
Db        1753 ------------------------------------GGCTACCGTGGACGTGGG--- 1770

Qy         247 AlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeuGlnGly 266
                                                              |||   :::|||
Db        1771 ------------------------------------------------GGATATCGCGGA 1782

Qy         267 ArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProProGlu 286
               |||          |||          |||   ||||||       |||
Db        1783 AGA----------GGAAGAGGCATGTACCATTATCACAACGACGTGCATAGGCCG--- 1827

Qy         287 GlnAspAspThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsn 306

Db        1827 ------------------------------------------------------------ 1827

Qy         307 AspProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArg------ 324
                                     ||||||         :::   ||||||
Db        1828 --------------------CCGCATCCTGGCTCCTTAGCAGGTTCCTCTTCTTCAGTT 1866

Qy         325 ProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValValGly 344
               |||      |||                                              |||
Db        1867 CCCCCTCCCGGC----------------------------------------------GGA 1881

Qy         345 LysGluProAsnTrpArgArgProSerAlaAsnGly------ThrGlyIleLeuProSer 362
                  |||      ||||||||||||                      |||      |||-
Db        1882 GTGGAGGATGAGTGGCGACGCCCGCCAGGAGAATCGTACGACAGAGGAGCCCGCTCCAGC 1941

Qy         363 ProThrGlyProGlyIle----------LeuProSerProAlaGlnGlyThrArgArg 378
                        ||||||:::              ||||||   :::      |||
Db        1942 TCTCGGGAACCTGGTGTAGAACGCTCTCGTTCTAGATCCCCATTGAAGCGGGCTCGCTCT 2001

Qy         379 Pro--------------------MetArgSerAsn---------------------- 383
               |||                     |||   |||
Db        2002 CCCGGATCTGATTCTGACACCTCGACTCGTCGGAACGACGCCCTTGCATCAGCCAGTACG 2061

Qy         384 ---ProAsp--------------------SerTrpGluGlyTyrAspProAlaGlnLeu 395
                  ||||||                     |||   |||         |||   :::
Db        2062 GTTCCAGATGTGGCGCGCAAGTGCAGCACAGTGTGGACCGGA---------GCGCTAATC 2112

Qy         396 ValArgGluSer------------LysArgThrArgArgAspGlySerValAspGlyPhe 411
               ::::::   |||             |||          ||||||       |||
Db        2113 CTCAAGAGCTCACTGTTCCCAGCCAAGTTTCATTTGACGGACGGAGATACGGATATTGTG 2172

Qy         412 ThrProMetGlyValAspGluArgSerPheGlyArgGlySerValAlaAlaArgProIle 431
                  :::      ||||||                        :::      |||
Db        2173 GAGTCGCTGATGCGCGATGAGGAGGGCAAGCACAACCTACGTATCACCCAGCGACTGCGT 2232

Qy         432 ArgGlyProProAspSerAspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrPro 451
                  ||||||      |||   :::   :::   :::   :::
Db        2233 CTGGATCCGCCCAAGCTGGACGACGTGCAAAAACGTATTGCATCCTCATCG--------- 2283

Qy         452 ValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluVal 471
```

```
                    :::      :::   ||||||           |||
Db       2284 ---------TCGCACGCCATTTTCATGGGTCTGGCAGGGTCCACA---------AACGAC 2325

Qy        472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIle 491
              ||||||
Db       2326 ACCAATTGC---------------------------------------------------- 2334

Qy        492 GlyCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGlu 511
                                                 :::|||   :::      |||
Db       2335 ----------------------------------GACGATGCCAGCGTGCAGACGCGA 2358

Qy        512 PheLeuArg------------TyrLeuSerSerLysAspArgAlaGlyValAlaLysLeu 527
              ||||||             ||||||      |||:::   |||||||||      |||
Db       2359 CCACTAAGAAACTTGGTTTCCTATCTTAAGCAGAAGGAGGCGGCTGGCGTGATCTCGCTG 2418

Qy        528 AspAspGlyThrThr--------------LeuPheLeuValProProSerAspPheLeu 542
                :::     |||               |||:::      ||||||   ||||||
Db       2419 CTTAACAAGGAGACGGAGGCAACTGGGGTGCTGTACGCCTTCCCCCCCTGCGACTTTTCC 2478

Qy        543 ThrAspValLeuGlnValThr-------------------------ArgGlnGluArg 553
              |||::::::|||:::   |||                        :::::::::
Db       2479 ACAGAGCTGCTAAAGCGCACTTGTCACAGTCTGACGGAGGAGGGCCTGAAGGAGGATCAC 2538

Qy        554 LeuTyrGlyValValLeuLys 560
              |||      ||||||::::::
Db       2539 CTGGTGATTGTGGTGGTGCGC 2559
```

```
RESULT 14
ABL05394/c
ID   ABL05394 standard; cDNA; 9744 BP.
XX
AC   ABL05394;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 10664.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB61291.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 10664; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 9744 BP; 2598 A; 2274 C; 2331 G; 2541 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:              1.24e-05        Length:         9744
Score:                  287.50          Matches:        140
Percent Similarity:     32.70%          Conservative:   65
Best Local Similarity:  22.33%          Mismatches:     201
Query Match:            5.99%           Indels:         221
DB:                     4               Gaps:           27

US-09-920-705-3 (1-901) x ABL05394 (1-9744)

Qy          11 AspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThr 30
               :::|||             :::      |||:::   |||:::|||
Db        7411 GAGGATGATCCTCTTTCTACGCGCACACTGTTTGCAGGAAATCTGGAGGTTACCATCGCC 7352

Qy          31 GluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAspArgIleThrValTyr 50
               :::   :::|||      :::||||||:::||||||   :::|||   |||   :::
Db        7351 GACGACGAACTGCGTCGCATCTTTGGCAAATACGGAGTGGTAGATGACATCGACATCAAG 7292

Qy          51 ---------------SerSerArgGlyPheAlaPheIleTyrTyrArgHisValGluGlu 65
                              :::         |||||||||:::   |||:::::::::::
Db        7291 CGACCGCCGCCGGGTACGGGGAACGCGTTCGCGTTTGTACGTTACCAAAACCTGGATATG 7232

Qy          66 AlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleLysIle 85
               |||      ||||||      |||   |||      :::          |||   ||||||
Db        7231 GCTCACCGAGCGAAGATCGAATTGTCTGGACAGTACATTGGCAAGTTTCAGTGCAAGATC 7172

Qy          86 GluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGlyGlyIleGlyProAsn 105
               |||   ::         |||          :::|||:::||||||:::|||
Db        7171 GGATACGGGAAA---GTGACTCCGGCAACTCGTATGTGGATCGGAGGCTTAGGTGCCTGG 7115

Qy         106 ValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArg 125
               |||         ||||||   ||||||   :::||||||   |||:::
Db        7114 ACTTCTGTCACTCAGCTGGAGCGAGAGTTTGATCGCTTCGGTGCGATCAAGAAGATTGAA 7055

Qy         126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
               :::   :::            |||:::|||   |||      ::::::   |||
Db        7054 TACCAGAAGGGCGAGCCGTACGCTTATATTCAGTACGAGACAGTGGAGGCAGCCACGGCG 6995

Qy         146 Ala---LysSerMetAsnGlyLysProMetGlyGly------SerPheLeuArgValAsp 162
               |||   |||   |||   |||   |||:::||||||            ||||||   |||
Db        6994 GCGGTAAAGGAGATGCGAGGATTCCCATTGGGTGGACCGGAGCGTCGACTCCGCACGGAC 6935

Qy         163 Phe--------------------------------LeuArgSerGlnAlaProLysLys 171
               |||                              :::|||      |||
Db        6934 TTTGCCGAGTTGCCCGGCGCCACACCGGCAGCGCCTTTCAAGAGCTCAAAGCCGCCATAT 6875

Qy         172 GluGlnTrpAlaGlySerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGlnTyr 191
               ::::::   |||      |||                          :::|||:::|||
Db        6874 GATGAAAGTGCGTTGGAGTAC----------------------AGACGTCCGGAATAC 6839

Qy         192 ProHisSerTyrGluAspPhe--------------LysGlyAspValGlnProSerLys 206
                     ||||||:::                  :::|||          |||
Db        6838 GATCCGTATTACGAAGAGTCCGCCGCCTACGCTCCGCGCGGTGGCTATTCTCCG------ 6785

Qy         207 ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHis 226
                           :::|||||||
Db        6784 --------------TATCCACCTAGGGGT---------------------------- 6770

Qy         227 AsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsnPhe 246
                                                          |||      |||
Db        6769 -------------------------------------GGCTACCGTGGACGTGGG--- 6752

Qy         247 AlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeuGlnGly 266
                                                                |||   :::|||
Db        6751 ----------------------------------------------GGATATCGCGGA 6740

Qy         267 ArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProProGlu 286
               |||              |||          |||   ||||||        |||
Db        6739 AGA------------GGAAGAGGCATGTACCATTATCACAACGACGTGCATAGGCCG--- 6695
```

```
Qy        287 GlnAspAspThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsn 306

Db       6695 ------------------------------------------------------------ 6695

Qy        307 AspProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArg------ 324

Db       6694 --------------------CCGCATCCTGGCTCCTTAGCAGGTTCCTCTTCTTCAGTT 6656

Qy        325 ProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValValGly 344

Db       6655 CCCCCTCCCGGC------------------------------------------GGA 6641

Qy        345 LysGluProAsnTrpArgArgProSerAlaAsnGly------ThrGlyIleLeuProSer 362

Db       6640 GTGGAGGATGAGTGGCGACGCCCGCCAGGAGAATCGTACGACAGAGGAGCCCGCTCCAGC 6581

Qy        363 ProThrGlyProGlyIle------------LeuProSerProAlaGlnGlyThrArgArg 378

Db       6580 TCTCGGGAACCTGGTGTAGAACGCTCTCGTTCTAGATCCCCATTGAAGCGGGCTCGCTCT 6521

Qy        379 Pro--------------------MetArgSerAsn----------------------- 383

Db       6520 CCCGGATCTGATTCTGACACCTCGACTCGTCGGAACGACGCCCTTGCATCAGCCAGTACG 6461

Qy        384 ---ProAsp--------------------SerTrpGluGlyTyrAspProAlaGlnLeu 395

Db       6460 GTTCCAGATGTGGCGCGCAAGTGCAGCACAGTGTGGACCGGA---------GCGCTAATC 6410

Qy        396 ValArgGluSer------------LysArgThrArgArgAspGlySerValAspGlyPhe 411

Db       6409 CTCAAGAGCTCACTGTTCCCAGCCAAGTTTCATTTGACGGACGGAGATACGGATATTGTG 6350

Qy        412 ThrProMetGlyValAspGluArgSerPheGlyArgGlySerValAlaAlaArgProIle 431

Db       6349 GAGTCGCTGATGCGCGATGAGGAGGGCAAGCACAACCTACGTATCACCCAGCGACTGCGT 6290

Qy        432 ArgGlyProProAspSerAspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrPro 451

Db       6289 CTGGATCCGCCCAAGCTGGACGACGTGCAAAAACGTATTGCATCCTCATCG--------- 6239

Qy        452 ValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluVal 471

Db       6238 ---------TCGCACGCCATTTTCATGGGTCTGGCAGGGTCCACA---------AACGAC 6197

Qy        472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIle 491

Db       6196 ACCAATTGC---------------------------------------------------- 6188

Qy        492 GlyCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGlu 511

Db       6187 ----------------------------------GACGATGCCAGCGTGCAGACGCGA 6164

Qy        512 PheLeuArg------------TyrLeuSerSerLysAspArgAlaGlyValAlaLysLeu 527

Db       6163 CCACTAAGAAACTTGGTTTCCTATCTTAAGCAGAAGGAGGCGGCTGGCGTGATCTCGCTG 6104

Qy        528 AspAspGlyThrThr--------------LeuPheLeuValProProSerAspPheLeu 542

Db       6103 CTTAACAAGGAGACGGAGGCAACTGGGGTGCTGTACGCCTTCCCCCCCTGCGACTTTTCC 6044

Qy        543 ThrAspValLeuGlnValThr--------------------------ArgGlnGluArg 553

Db       6043 ACAGAGCTGCTAAAGCGCACTTGTCACAGTCTGACGGAGGAGGGCCTGAAGGAGGATCAC 5984

Qy        554 LeuTyrGlyValValLeuLys 560

Db       5983 CTGGTGATTGTGGTGGTGCGC 5963
```

RESULT 15
ABZ21697
ID ABZ21697 standard; cDNA; 3383 BP.
XX
AC ABZ21697;
XX
DT 27-FEB-2003 (first entry)
XX
DE Human RBM15-L protein encoding cDNA SEQ ID NO:11.
XX
KW Human; RBM15; RNA binding motif protein 15; megakaryoblastic leukaemia 1;
KW MKL1; fusion protein; acute megakaryoblastic leukaemia; AMKL; cytostatic;
KW t(1; 22) chromosomal rearrangement; gene therapy; chromosome 1p13; gene;
KW ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 84. .2957
FT /*tag= a
FT /product= "RBM15-L protein"
XX
PN WO200288309-A2.
XX
PD 07-NOV-2002.
XX
PF 23-APR-2002; 2002WO-US012797.
XX
PR 27-APR-2001; 2001US-0286910P.
XX
PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PA (HOSP-) HOSPITAL FOR SICK CHILDREN.
XX
PI Morris SW, Ma Z, Hitzler JK;
XX
DR WPI; 2003-103455/09.
DR P-PSDB; ABP56070.
XX
PT New RNA-binding motif protein-15 (RBM15)-megakaryoblastic leukemia-1
PT (MKL1), MKL1-RBM15-S and MKL1-RBM15-S+AE fusion proteins, useful for
PT identifying agents useful for treating patients with acute
PT megakaryoblastic leukemia.
XX
PS Disclosure; Page 89-93; 109pp; English.
XX
CC The present invention describes an RNA-binding motif protein-15 (RBM15)-
CC megakaryoblastic leukaemia-1 (MKL1) fusion protein, a MKL1-RBM15-S fusion
CC protein, and a MKL1-RBM15-S+AE fusion protein associated with acute
CC megakaryoblastic leukaemia (AMKL). Also described: (1) an antibody that
CC specifically binds to the RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE
CC fusion proteins; (2) a non-human transgenic animal that has been altered
CC to express a gene encoding a RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE
CC fusion protein; (3) identifying an agent capable of binding to a RBM15-
CC MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE fusion protein; (4) detecting the
CC t(1;22) chromosomal rearrangement associated with AMKL; and (5) screening
CC for agents capable of (selectively) inhibiting the activity of a fusion
CC protein arising from the t(1;22) chromosomal rearrangement associated
CC with AMKL. The fusion proteins have cytostatic activity and can be used
CC in gene therapy. The RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE fusion
CC proteins and nucleotide molecules are useful for designing and preparing
CC agents that specifically inhibit the expression of the RBM15-MKL1 or MKL1
CC -RBM15 genes in cells for therapeutic and other purposes. The transgenic
CC animals are useful for identifying and testing carcinogenic or
CC therapeutic compositions. The methods are also useful for detecting the
CC t(1;22) chromosomal rearrangement associated with AMKL, or for
CC identifying agents useful for treating patients with AMKL. The antibodies
CC can be used to selectively kill cells expressing RBM15-MKL1, MKL1-RBM15-
CC S, or MKL1-RBM15-S+AE fusion protein. RBM15 is located to chromosome
CC 1p13, and MKL1 is located to chromosome 22q13. The present sequence
CC encodes the human RBM15-L protein which is given in the exemplification
CC of the present invention
XX
SQ Sequence 3383 BP; 812 A; 854 C; 952 G; 765 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.53e-05	Length:	3383
Score:	276.50	Matches:	162

Percent Similarity: 35.87% Conservative: 88
Best Local Similarity: 23.24% Mismatches: 256
Query Match: 5.76% Indels: 191
DB: 8 Gaps: 34

US-09-920-705-3 (1-901) x ABZ21697 (1-3383)

```
Qy          7 ProPheArg-----------------AlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
Db       1152 CCTTTCAGAGAAGTGGATGAGATTTCACCCGAGGATGATCAGCGAGCTAACCGGACGCTC 1211

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
Db       1212 TTCTTGGGCAACCTAGACATCACTGTAACGGAGAGTGATTTAAGAAGGGCGTTTGATCGC 1271

Qy         41 TyrGlyAspIleAspArgIleThrVal---TyrSerSerArgGly-----------Phe 55
Db       1272 TTTGGAGTCATCACAGAAGTAGATATCAAGAGGCCTTCTCGCGGCCAGACTAGTACTTAC 1331

Qy         56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGly 75
Db       1332 GGCTTTCTCAAATTTGAGAACTTAGATATGTCTCACCGGGCCAAATTAGCAATGTCTGGC 1391

Qy         76 AlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLys 95
Db       1392 AAAATTATAATTCGGAATCCTATCAAAATTGGTTATGGTAAA---GCTACACCCACCACC 1448

Qy         96 SerLeuTrpValGlyGlyIleGlyProAsnValSerLysAspAspLeuGluGluGluPhe 115
Db       1449 CGCCTCTGGGTGGGAGGCCTGGGACCTTGGGTTCCTCTTGCTGCCCTGGCACGAGAATTT 1508

Qy        116 SerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIle 135
Db       1509 GATCGATTTGGCACCATACGCACCATAGACTACCGAAAAGGTGATAGTTGGGCATATATC 1568

Qy        136 AspTyrTyrGluMetAspAspAlaLeuGlnAla---LysSerMetAsnGlyLysProMet 154
Db       1569 CAGTATGAAAGCCTGGATGCAGCGCATGCTGCCTGGACCCATATGCGGGGCTTCCCACTT 1628

Qy        155 GlyGly------SerPheLeuArgValAspPheLeuArgSerGlnAlaProLysLysGlu 172
Db       1629 GGTGGCCCAGATCGACGCCTTAGAGTAGACTTTGCCGACACCGAACATCGTTACCAGCAG 1688

Qy        173 GlnTrp-----------------------------------AlaGlySerTyrAspAsn 180
Db       1689 CAGTATCTGCAGCCTCTGCCCTTGACTCATTATGAGCTGGTGACAGATGCTTTTGGACAT 1748

Qy        181 Arg--------------AsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyr 195
Db       1749 CGGGCACCAGACCCTTTGAGGGGTGCTCGGGATAGGACACCA------CCCTTACTATAC 1802

Qy        196 GluAspPheLysGlyAspValGlnProSerLysValLeuTrpIleGlyPheProProThr 215
Db       1803 AGAGATCGTGATAGGGACCTTTATCCTGACTCTGAT---TGGGTGCCACCCCCACCCCCA 1859

Qy        216 AlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIle 235
Db       1860 GTCCGA-----------GAACGCAGCACTCGGACTGCAGCTACTTCTGTGCCTGCTTAT 1907

Qy        236 GluArgValLysSerTyrProSerArg------AsnPheAlaLeuValGluPheArgSer 253
Db       1908 GAGCCACTGGATAGCCTAGATCGCAGGCGGGATGGTTGGTCCTTGGACCGGGACAGAGGT 1967

Qy        254 AlaGluGluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArg 273
Db       1968 GATCGAGATCTGCCCAGCAGCAGAGAC----------------------CAGCCTAGG 2003

Qy        274 IleLysIleMetTyrSerAsnAspGluLeuProProGluGlnAsp-----------Asp 289
Db       2004 AAGCGA----------------AGGCTGCCTGAGGAGAGTGGAGGACGTCATCTGGAT 2045

Qy        290 ThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsnAspProSer 309
```

```
Db       2046 AGGTCTCCTGAGAGTGACCGCCCACGAAAACGT----------------------CAC 2081

Qy        310 CysValSerSerPro---HisSerThrGlyIleProGlySerMetArgProLeuArgGly 328
Db       2082 TGCGCTCCTTCTCCTGACCGCAGTCCAGAATTGAGCAGTAGCCGGGATCGTTACAACAGC 2141

Qy        329 ThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValValGlyLysGluProAsn 348
Db       2142 GACAATGATCGATCT----------------TCCCGTCTTCTCTTGGAAAGGCCCTCT 2183

Qy        349 TrpArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIle 368
Db       2184 ---------CCAATCAGAGACGGACGAGGTAGTTTG---------------------- 2210

Qy        369 LeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrpGlu 388
Db       2211 ------GAGAAGAGCCAGGGTGACAAGCGAGAC---CGTAAAAACTCTGCATCAGCTGAA 2261

Qy        389 GlyTyrAspProAlaGlnLeuValArgGluSerLysArgThrArgArgAspGlySerVal 408
Db       2262 -------------------------CGAGATAGGAAGCACCGGACAACTGCTCCCACT 2294

Qy        409 AspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGly---------Ser 425
Db       2295 GAGGGAAAAAGCCCTCTGAAAAAAGAAGACCGCTCTGATGGGAGTGCACCTAGCACCAGC 2354

Qy        426 ValAlaAlaArgProIleArgGlyPro---------------------------------- 434
Db       2355 ACTGCTTCCTCCAAGCTGAAGTCCCCGTCCCAGAAACAGGATGGGGGGACAGCCCCTGTG 2414

Qy        435 ------------ProAspSerAspHisIleTrpArgGlyMet------------------ 444
Db       2415 GCATCAGCCTCTCCCAAACTCTGTTTGGCCTGGCAGGGCATGCTTCTACTGAAGAACAGC 2474

Qy        445 ---------------------IleAlaLysGlyGlyThrProValCysCysAlaArgCys 457
Db       2475 AACTTTCCTTCCAACATGCATCTGTTGCAGGGTGACCTCCAAGTGGCTAGTAGTCTTCTT 2534

Qy        458 Val---------ProMetGlyLysGlyIleGluThrLysLeuProGluValValAsnCys 474
Db       2535 GTGGAGGGTTCAACTGGAGGCAAAGTGGCCCAGCTCAAGATCACTCAGCGTCTCCGTTTG 2594

Qy        475 SerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAla------IleGly 492
Db       2595 GAC---CAGCCCAAGTTGGATGAAGTAACTCGACGCATCAAAGTAGCAGGGCCCAATGGT 2651

Qy        493 CysGluIleValPhePheValProAspArgGluGlu----------------------- 504
Db       2652 TATGCCATTCTTTTGGCTGTGCCTGGAAGTTCTGACAGCCGGTCCTCCTCTTCCTCAGCT 2711

Qy        505 ------AspPheAlaSerTyrThrGlu--------------PheLeuArgTyrLeuSer 517
Db       2712 GCATCAGACACTGCCACTTCTACTCAGAGGCCACTTAGGAACCTTGTGTCCTATTTAAAG 2771

Qy        518 SerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValPro 537
Db       2772 CAAAAGCAGGCAGCCGGGGTGATCAGCCTCCCTGTGGGGGGC----------------- 2813

Qy        538 ProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyVal 557
Db       2814 --------------------------------AACAAAGACAAGGAAAACACCGGGGTC 2840

Qy        558 ValLeuLysLeuProProProAlaValProValThrAlaSerTyrArgGlnGluSerGln 577
Db       2841 CTTCATGCCTTCCCA---------------------------------------------- 2855

Qy        578 SerAsnProLeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSer 597
Db       2856 ------CCTTGTGAGTTCTCCCAGCAGTTCCTGGATTCCCCTGCCAAGGCA--------- 2900

Qy        598 LeuTyrProProArgGluAsnTyr--------------IleArgGlyAla 609
Db       2901 CTGGCCAAATCTGAAGAAGATTACCTGGTCATGATCATTGTCCGTGGTGCG 2951
```

Search completed: April 8, 2005, 09:07:22
Job time : 1324 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2005, 08:37:33 ; Search time 370 Seconds
(without alignments)
3984.555 Million cell updates/sec

Title: US-09-920-705-3
Perfect score: 4801
Sequence: 1 MALSMKPFRADDSGFQSNNL.....LQIQQKQQQQSSGTPAGQGP 901

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09920705/runat_07042005_101814_20701/app_query.fasta_1.1095
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09920705_@CGN_1_1_75_@runat_07042005_101814_20701 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	271.5	5.7	11231	4	US-09-949-016-14030	Sequence 14030, A
2	268	5.6	3192	4	US-09-949-016-2288	Sequence 2288, Ap
3	268	5.6	3228	4	US-09-799-451-752	Sequence 752, App
4	246.5	5.1	1477	4	US-09-976-594-794	Sequence 794, App
5	246	5.1	2228	1	US-07-726-607C-1	Sequence 1, Appli
6	246	5.1	2228	1	US-07-843-949A-1	Sequence 1, Appli
7	246	5.1	2228	2	US-08-218-978-1	Sequence 1, Appli
8	246	5.1	2356	4	US-09-949-016-4865	Sequence 4865, Ap
9	246	5.1	2356	4	US-09-949-016-4866	Sequence 4866, Ap
10	242.5	5.1	1401	1	US-07-843-949A-3	Sequence 3, Appli
11	242.5	5.1	1401	2	US-08-218-978-3	Sequence 3, Appli
12	236.5	4.9	8147	3	US-09-514-247A-9	Sequence 9, Appli
13	215.5	4.5	1968	4	US-09-949-016-3886	Sequence 3886, Ap
14	208	4.3	4674	4	US-09-410-551B-26	Sequence 26, Appl
15	208	4.3	4674	4	US-09-940-316B-26	Sequence 26, Appl
16	207	4.3	1323	4	US-09-949-016-4868	Sequence 4868, Ap
17	207	4.3	1323	4	US-09-949-016-4869	Sequence 4869, Ap
18	201.5	4.2	4737	4	US-09-410-551B-30	Sequence 30, Appl
19	201.5	4.2	4737	4	US-09-940-316B-30	Sequence 30, Appl
20	200.5	4.2	1467	1	US-07-881-075-50	Sequence 50, Appl
21	200.5	4.2	1467	1	US-08-120-827-50	Sequence 50, Appl
22	200.5	4.2	1467	1	US-08-478-675-50	Sequence 50, Appl
23	200.5	4.2	1469	1	US-08-461-027-1	Sequence 1, Appli
24	200.5	4.2	1469	3	US-09-082-501-1	Sequence 1, Appli
25	200.5	4.2	1469	5	PCT-US92-01836-1	Sequence 1, Appli
26	200.5	4.2	2559	4	US-09-220-132-135	Sequence 135, App
27	200.5	4.2	2562	1	US-08-146-421-4	Sequence 4, Appli
c 28	200.5	4.2	13987	2	US-08-804-227C-13	Sequence 13, Appl
c 29	200.5	4.2	44377	2	US-08-804-227C-7	Sequence 7, Appli
c 30	200.5	4.2	44377	2	US-08-804-198-1	Sequence 1, Appli
31	200	4.2	2642	4	US-09-949-016-1282	Sequence 1282, Ap
32	199.5	4.2	2731	4	US-09-976-594-281	Sequence 281, App
33	198.5	4.1	9046	1	US-08-227-536-1	Sequence 1, Appli
34	198.5	4.1	9046	5	PCT-US95-04682-1	Sequence 1, Appli
35	195.5	4.1	1231	4	US-09-949-016-1766	Sequence 1766, Ap
36	195.5	4.1	1416	3	US-08-866-928B-2	Sequence 2, Appli
37	195.5	4.1	1416	4	US-09-685-836-2	Sequence 2, Appli
38	195.5	4.1	4767	4	US-09-410-551B-28	Sequence 28, Appl
39	195.5	4.1	4767	4	US-09-940-316B-28	Sequence 28, Appl
40	195.5	4.1	4818	4	US-09-410-551B-32	Sequence 32, Appl
41	195.5	4.1	4818	4	US-09-940-316B-32	Sequence 32, Appl
42	194.5	4.1	2277	1	US-08-676-967-5	Sequence 5, Appli
43	194.5	4.1	2277	1	US-08-676-974-5	Sequence 5, Appli
44	194.5	4.1	2277	2	US-09-098-487-5	Sequence 5, Appli
45	194	4.0	4725	4	US-09-410-551B-24	Sequence 24, Appl

ALIGNMENTS

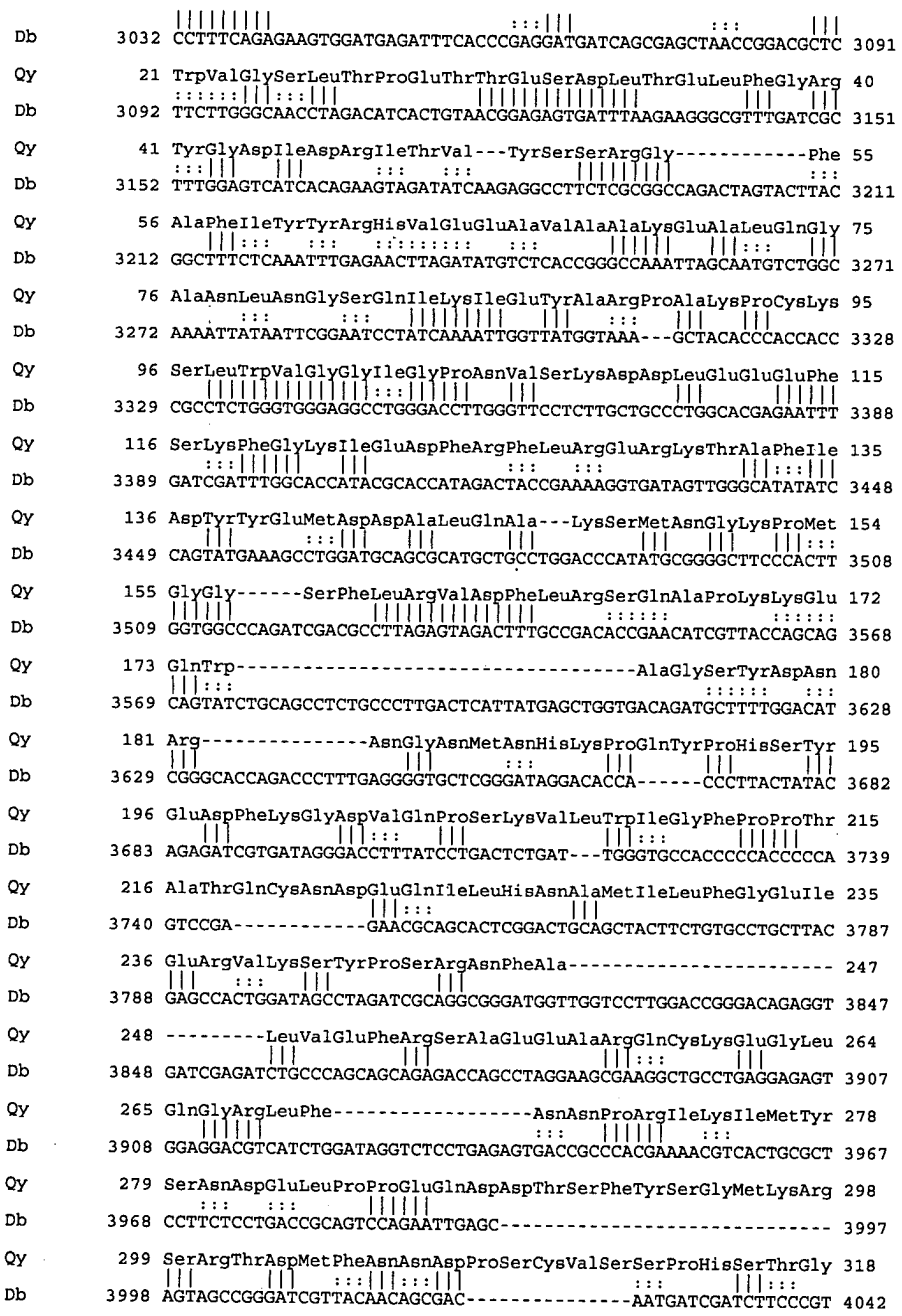
RESULT 1
US-09-949-016-14030
; Sequence 14030, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14030
; LENGTH: 11231
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14030

Alignment Scores:

Pred. No.:	2.24e-11	Length:	11231
Score:	271.50	Matches:	158
Percent Similarity:	34.32%	Conservative:	86
Best Local Similarity:	22.22%	Mismatches:	248
Query Match:	5.66%	Indels:	219
DB:	4	Gaps:	31

US-09-920-705-3 (1-901) x US-09-949-016-14030 (1-11231)

Qy 7 ProPheArg----------AlaAspAspSerGlyPheGlnSerAsnAsnLeu 20



RESULT 2
US-09-949-016-2288
; Sequence 2288, Application US/09949016

```
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2288
;   LENGTH: 3192
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-2288

Alignment Scores:
Pred. No.:                6.07e-12        Length:          3192
Score:                    268.00          Matches:         155
Percent Similarity:       34.28%          Conservative:    86
Best Local Similarity:    22.05%          Mismatches:      248
Query Match:              5.58%           Indels:          214
DB:                       4               Gaps:            30

US-09-920-705-3 (1-901) x US-09-949-016-2288 (1-3192)

Qy          7 ProPheArg-----------------AlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              |||||||||                 :::|||           :::      |||
Db       1032 CCTTTCAGAGAAGTGGATGAGATTTCACCCGAGGATGATCAGCGAGCTAACCGGACGCTC 1091

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ::::::|||:::|||           |||||||||||||||          |||   |||
Db       1092 TTCTTGGGCAACCTAGACATCACTGTAACGGAGAGTGATTTAAGAAGGGCGTTTGATCGC 1151

Qy         41 TyrGlyAspIleAspArgIleThrVal---TyrSerSerArgGly-----------Phe 55
              :::|||   |||     :::  :::       |||||||||           :::
Db       1152 TTTGGAGTCATCACAGAAGTAGATATCAAGAGGCCTTCTCGCGGCCAGACTAGTACTTAC 1211

Qy         56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGly 75
                 |||:::   :::   :::::::::   :::      ||||||   |||:::   |||
Db       1212 GGCTTTCTCAAATTTGAGAACTTAGATATGTCTCACCGGGCCAAATTAGCAATGTCTGGC 1271

Qy         76 AlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLys 95
                 :::      :::   |||||||||   |||   :::   |||   |||
Db       1272 AAAATTATAATTCGGAATCCTATCAAAATTGGTTATGGTAAA---GCTACACCCACCACC 1328

Qy         96 SerLeuTrpValGlyGlyIleGlyProAsnValSerLysAspAspLeuGluGluGluPhe 115
              |||||||||||||||:::||||||   |||       |||      ||||||
Db       1329 CGCCTCTGGGTGGGAGGCCTGGGACCTTGGGTTCCTCTTGCTGCCCTGGCACGAGAATTT 1388

Qy        116 SerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIle 135
                ::||||||   |||           :::   :::           |||:::|||
Db       1389 GATCGATTTGGCACCATACGCACCATAGACTACCGAAAAGGTGATAGTTGGGCATATATC 1448

Qy        136 AspTyrTyrGluMetAspAspAlaLeuGlnAla---LysSerMetAsnGlyLysProMet 154
                 |||      ::|||   |||       |||         |||   |||   |||:::
Db       1449 CAGTATGAAAGCCTGGATGCAGCGCATGCTGCCTGGACCCATATGCGGGGCTTCCCACTT 1508

Qy        155 GlyGly------SerPheLeuArgValAspPheLeuArgSerGlnAlaProLysLysGlu 172
              ||||||           ||||||||||||||||      ::::::          ::::::
Db       1509 GGTGGCCCAGATCGACGCCTTAGAGTAGACTTTGCCGACACCGAACATCGTTACCAGCAG 1568

Qy        173 GlnTrp-----------------------------------AlaGlySerTyrAspAsn 180
              |||:::                                   ::::::   :::
Db       1569 CAGTATCTGCAGCCTCTGCCCTTGACTCATTATGAGCTGGTGACAGATGCTTTTGGACAT 1628

Qy        181 Arg---------------AsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyr 195
                 |||           |||      |||   :::      |||      |||
Db       1629 CGGGCACCAGACCCTTTGAGGGGTGCTCGGGATAGGACACCA------CCCTTACTATAC 1682

Qy        196 GluAspPheLysGlyAspValGlnProSerLysValLeuTrpIleGlyPheProProThr 215
                 |||          |||:::  |||          |||:::      ||||||
Db       1683 AGAGATCGTGATAGGGACCTTTATCCTGACTCTGAT---TGGGTGCCACCCCCACCCCCA 1739

Qy        216 AlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIle 235
                          |||:::          |||
Db       1740 GTCCGA-----------GAACGCAGCACTCGGACTGCAGCTACTTCTGTGCCTGCTTAC 1787

Qy        236 GluArgValLysSerTyrProSerArgAsnPheAla---------------------- 247
                 |||   :::   |||       |||
Db       1788 GAGCCACTGGATAGCCTAGATCGCAGGCGGGATGGTTGGTCCTTGGACCGGGACAGAGGT 1847

Qy        248 ---------LeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeu 264
                       |||       |||          |||:::      |||
Db       1848 GATCGAGATCTGCCCAGCAGCAGAGACCAGCCTAGGAAGCGAAGGCTGCCTGAGGAGAGT 1907

Qy        265 GlnGlyArgLeuPhe-----------------AsnAsnProArgIleLysIleMetTyr 278
                 ||||||                    :::   |||||   :::
Db       1908 GGAGGACGTCATCTGGATAGGTCTCCTGAGAGTGACCGCCCACGAAAACGTCACTGCGCT 1967

Qy        279 SerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArg 298
                 :::   :::      ||||||
Db       1968 CCTTCTCCTGACCGCAGTCCAGAATTGAGC---------------------------- 1997

Qy        299 SerArgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGly 318
              |||      |||   :::|||:::|||           :::      |||:::
Db       1998 AGTAGCCGGGATCGTTACAACAGCGAC--------------AATGATCGATCTTCCCGT 2042

Qy        319 Ile------------ProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyr 334
                 :::          |||     :::|||     |||||:::   |||:::|||
Db       2043 CTTCTCTTGGAAAGGCCCTCTCCAATCAGAGACAGACGAGGTAGTTTGGAGAAGAGC--- 2099

Qy        335 AsnGlyAlaGluTyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAla 354

Db       2099 ------------------------------------------------------------ 2099

Qy        355 AsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGln 374
                                                                      |||
Db       2100 ----------------------------------------------------------CAG 2102

Qy        375 GlyThrArgArgProMetArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGln 394
              |||   :::|||    |||  |||     |||  |||
Db       2103 GGTGACAAGCGAGAC---CGTAAAAACTCTGCATCAGCTGAA----------------- 2141

Qy        395 LeuValArgGluSerLysArgThrArgArgAspGlySerValAspGlyPheThrProMet 414
                       ::::::   |||               :::|||   :::|||:::
Db       2142 ---------CGAGATAGGAAGCACCGGACAACTGCTCCCACTGAGGGAAAAAGCCCTCTG 2192

Qy        415 GlyValAspGluArgSerPheGlyArgGly---------SerValAlaAlaArgProIle 431
                    ::::::||||||   |||          |||   |||:::      :::
Db       2193 AAAAAAGAAGACCGCTCTGATGGGAGTGCACCTAGCACCAGCACTGCTTCCTCCAAGCTG 2252

Qy        432 ArgGlyPro-----------------------------------------------ProAsp 436
                 :::   |||                                           |||
Db       2253 AAGTCCCCGTCCCAGAAACAGGATGGGGGGACAGCCCCTGTGGCATCAGCCTCTCCCAAA 2312

Qy        437 SerAspHisIleTrpArgGlyMet---------------------------------- 444
                       |||:::||||||
Db       2313 CTCTGTTTGGCCTGGCAGGGCATGCTTCTACTGAAGAACAGCAACTTTCCTTCCAACATG 2372

Qy        445 ---IleAlaLysGlyGlyThrProValCysCysAlaArgCysVal---------ProMet 460
                    :::   :::|||          |||      :::      |||
Db       2373 CATCTGTTGCAGGGTGACCTCCAAGTGGCTAGTAGTCTTCTTGTGGAGGGTTCAACTGGA 2432

Qy        461 GlyLysGlyIleGluThrLysLeuProGluValValAsnCysSerAlaArgThrAspLeu 480
              ||||||      :::   |||:::  :::  :::      :::           |||
Db       2433 GGCAAAGTGGCCCAGCTCAAGATCACTCAGCGTCTCCGTTTGGAC---CAGCCCAAGTTG 2489

Qy        481 AsnMetLeuAlaLysHisTyrAlaValAla------IleGlyCysGluIleValPhePhe 498
                 :::   :::   :::         ||||||         |||      |||:::
```

```
Db       2490 GATGAAGTAACTCGACGCATCAAAGTAGCAGGGCCCAATGGTTATGCCATTCTTTTGGCT 2549

Qy        499 ValProAspArgGluGlu----------------------------AspPheAlaSer 508
               ||||||      :::                              |||   |||:::
Db       2550 GTGCCTGGAAGTTCTGACAGCCGGTCCTCCTCTTCCTCAGCTGCATCAGACACTGCCACT 2609

Qy        509 TyrThrGlu--------------PheLeuArgTyrLeuSerSerLysAspArgAlaGly 523
                 |||:::              :::   ||||||      |||     ||||||
Db       2610 TCTACTCAGAGGCCACTTAGGAACCTTGTGTCCTATTTAAAGCAAAAGCAGGCAGCCGGG 2669

Qy        524 ValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProProSerAspPheLeuThr 543
               |||     |||     |||
Db       2670 GTGATCAGCCTCCCTGTGGGGGGC---------------------------------- 2693

Qy        544 AspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLysLeuProPro 563
                         :::   :::         ||||||:::       |||
Db       2694 --------------AACAAAGACAAGGAAAACACCGGGGTCCTTCATGCCTTCCCA--- 2735

Qy        564 ProAlaValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnProLeuHisTyr 583
                                                           |||      :::
Db       2736 ------------------------------------------------CCTTGTGAGTTC 2747

Qy        584 MetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGlu 603
                  |||     ||||||||||||   |||         |||              |||
Db       2748 TCCCAGCAGTTCCTGGATTCCCCTGCCAAGGCA---------CTGGCCAAATCTGAAGAA 2798

Qy        604 AsnTyrIle 606
               :::|||:::
Db       2799 GATTACCTG 2807
```

RESULT 3
US-09-799-451-752
; Sequence 752, Application US/09799451
; Patent No. 6783969
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Ma, Yunqing
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Dunrui
; APPLICANT: Yang, Yonghong
; APPLICANT: Wehrman, Tom
; APPLICANT: Ghosh, Reena
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6783969el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 803
; CURRENT APPLICATION NUMBER: US/09/799,451
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 948
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 752
; LENGTH: 3228
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (108)..(2882)
US-09-799-451-752

Alignment Scores:

Pred. No.:	6.18e-12	Length:	3228
Score:	268.00	Matches:	155
Percent Similarity:	34.28%	Conservative:	86
Best Local Similarity:	22.05%	Mismatches:	248
Query Match:	5.58%	Indels:	214
DB:	4	Gaps:	30

US-09-920-705-3 (1-901) x US-09-799-451-752 (1-3228)

```
Qy          7 ProPheArg-----------------AlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              |||||||||                    :::|||          :::      |||
Db       1044 CCTTTCAGAGAAGTGGATGAGATTTCACCCGAGGATGATCAGCGAGCTAACCGGACGCTC 1103

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              :::::::|||:::|||              ||||||||||||||||          |||   |||
Db       1104 TTCTTGGGCAACCTAGACATCACTGTAACGGAGAGTGATTTAAGAAGGGCGTTTGATCGC 1163

Qy         41 TyrGlyAspIleAspArgIleThrVal---TyrSerSerArgGly------------Phe 55
              :::|||  |||      :::  :::        ||||||||                :::
Db       1164 TTTGGAGTCATCACAGAAGTAGATATCAAGAGGCCTTCTCGCGGCCAGACTAGTACTTAC 1223

Qy         56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGly 75
                 |||:::   :::   :::::::::   :::       ||||||   |||:::   |||
Db       1224 GGCTTTCTCAAATTTGAGAACTTAGATATGTCTCACCGGGCCAAATTAGCAATGTCTGGC 1283

Qy         76 AlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLys 95
                    :::      :::   |||||||||   |||   :::   |||   |||
Db       1284 AAAATTATAATTCGGAATCCTATCAAAATTGGTTATGGTAAA---GCAACACCCACCACC 1340

Qy         96 SerLeuTrpValGlyGlyIleGlyProAsnValSerLysAspAspLeuGluGluGluPhe 115
              |||||||||||||||:::|||||   |||             |||       ||||||
Db       1341 CGCCTCTGGGTGGGAGGCCTGGGACCTTGGGTTCCTCTTGCTGCCCTGGCACGAGAATTT 1400

Qy        116 SerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIle 135
               :::|||||   |||          :::  :::          |||:::|||
Db       1401 GATCGATTTGGCACCATACGCACCATAGACTACCGAAAAGGTGATAGTTGGGCATATATC 1460

Qy        136 AspTyrTyrGluMetAspAspAlaLeuGlnAla---LysSerMetAsnGlyLysProMet 154
               |||     :::|||  |||      |||       |||  |||  |||:::
Db       1461 CAGTATGAAAGCCTGGATGCAGCGCATGCTGCCTGGACCCATATGCGGGGCTTCCCACTT 1520

Qy        155 GlyGly------SerPheLeuArgValAspPheLeuArgSerGlnAlaProLysLysGlu 172
              ||||||              ||||||||||||||||      ::::::        ::::::
Db       1521 GGTGGCCCAGATCGACGCCTTAGAGTAGACTTTGCCGACACCGAACATCGTTACCAGCAG 1580

Qy        173 GlnTrp-----------------------------------AlaGlySerTyrAspAsn 180
              |||:::                                   ::::::   :::
Db       1581 CAGTATCTGCAGCCTCTGCCCTTGACTCATTATGAGCTGGTGACAGATGCTTTTGGACAT 1640

Qy        181 Arg--------------AsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyr 195
              |||                |||   :::   |||   |||      |||
Db       1641 CGGGCACCAGACCCTTTGAGGGGTGCTCGGGATAGGACACCA------CCCTTACTATAC 1694

Qy        196 GluAspPheLysGlyAspValGlnProSerLysValLeuTrpIleGlyPheProProThr 215
               |||     |||:::   |||          |||:::      ||||||
Db       1695 AGAGATCGTGATAGGGACCTTTATCCTGACTCTGAT---TGGGTGCCACCCCCACCCCCA 1751

Qy        216 AlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIle 235
                              |||:::      |||
Db       1752 GTCCGA------------GAACGCAGCACTCGGACTGCAGCTACTTCTGTGCCTGCTTAC 1799

Qy        236 GluArgValLysSerTyrProSerArgAsnPheAla---------------------- 247
               |||   :::   |||         |||
Db       1800 GAGCCACTGGATAGCCTAGATCGCAGGCGGGATGGTTGGTCCTTGGACCGGGACAGAGGT 1859

Qy        248 ---------LeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeu 264
                      |||         |||         |||:::      |||
Db       1860 GATCGAGATCTGCCCAGCAGCAGAGACCAGCCTAGGAAGCGAAGGCTGCCTGAGGAGAGT 1919

Qy        265 GlnGlyArgLeuPhe-----------------AsnAsnProArgIleLysIleMetTyr 278
                    ||||||                 :::   ||||||  :::
Db       1920 GGAGGACGTCATCTGGATAGGTCTCCTGAGAGTGACCGCCCACGAAAACGTCACTGCGCT 1979

Qy        279 SerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArg 298
                  :::   :::      ||||||
Db       1980 CCTTCTCCTGACCGCAGTCCAGAATTGAGC---------------------------- 2009
```

```
Qy       299 SerArgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGly 318
Db      2010 AGTAGCCGGGATCGTTACAACAGCGAC--------------AATGATCGATCTTCCCGT 2054

Qy       319 Ile------------ProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyr 334
Db      2055 CTTCTCTTGGAAAGGCCCTCTCCAATCAGAGACAGACGAGGTAGTTTGGAGAAGAGC--- 2111

Qy       335 AsnGlyAlaGluTyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAla 354
Db      2111 ------------------------------------------------------------ 2111

Qy       355 AsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGln 374
Db      2112 ----------------------------------------------------------CAG 2114

Qy       375 GlyThrArgArgProMetArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGln 394
Db      2115 GGTGACAAGCGAGAC---CGTAAAAACTCTGCATCAGCTGAA------------------- 2153

Qy       395 LeuValArgGluSerLysArgThrArgArgAspGlySerValAspGlyPheThrProMet 414
Db      2154 ---------CGAGATAGGAAGCACCGGACAACTGCTCCCACTGAGGGAAAAAGCCCTCTG 2204

Qy       415 GlyValAspGluArgSerPheGlyArgGly---------SerValAlaAlaArgProIle 431
Db      2205 AAAAAAGAAGACCGCTCTGATGGGAGTGCACCTAGCACCAGCACTGCTTCCTCCAAGCTG 2264

Qy       432 ArgGlyPro----------------------------------------------ProAsp 436
Db      2265 AAGTCCCCGTCCCAGAAACAGGATGGGGGGACAGCCCCTGTGGCATCAGCCTCTCCCAAA 2324

Qy       437 SerAspHisIleTrpArgGlyMet------------------------------------ 444
Db      2325 CTCTGTTTGGCCTGGCAGGGCATGCTTCTACTGAAGAACAGCAACTTTCCTTCCAACATG 2384

Qy       445 ---IleAlaLysGlyGlyThrProValCysCysAlaArgCysVal---------ProMet 460
Db      2385 CATCTGTTGCAGGGTGACCTCCAAGTGGCTAGTAGTCTTCTTGTGGAGGGTTCAACTGGA 2444

Qy       461 GlyLysGlyIleGluThrLysLeuProGluValValAsnCysSerAlaArgThrAspLeu 480
Db      2445 GGCAAAGTGGCCCAGCTCAAGATCACTCAGCGTCTCCGTTTGGAC---CAGCCCAAGTTG 2501

Qy       481 AsnMetLeuAlaLysHisTyrAlaValAla------IleGlyCysGluIleValPhePhe 498
Db      2502 GATGAAGTAACTCGACGCATCAAAGTAGCAGGGCCCAATGGTTATGCCATTCTTTTGGCT 2561

Qy       499 ValProAspArgGluGlu------------------------------AspPheAlaSer 508
Db      2562 GTGCCTGGAAGTTCTGACAGCCGGTCCTCCTCTTCCTCAGCTGCATCAGACACTGCCACT 2621

Qy       509 TyrThrGlu---------------PheLeuArgTyrLeuSerSerLysAspArgAlaGly 523
Db      2622 TCTACTCAGAGGCCACTTAGGAACCTTGTGTCCTATTTAAAGCAAAAGCAGGCAGCCGGG 2681

Qy       524 ValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProProSerAspPheLeuThr 543
Db      2682 GTGATCAGCCTCCCTGTGGGGGGC---------------------------------- 2705

Qy       544 AspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLysLeuProPro 563
Db      2706 ---------------AACAAAGACAAGGAAAACACCGGGGTCCTTCATGCCTTCCCA--- 2747

Qy       564 ProAlaValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnProLeuHisTyr 583
Db      2748 ------------------------------------------------CCTTGTGAGTTC 2759

Qy       584 MetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGlu 603
Db      2760 TCCCAGCAGTTCCTGGATTCCCCTGCCAAGGCA---------CTGGCCAAATCTGAAGAA 2810
```

```
Qy       604 AsnTyrIle 606
Db      2811 GATTACCTG 2819

RESULT 4
US-09-976-594-794
; Sequence 794, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
;  APPLICANT: Furness, Michael
;  APPLICANT:  Buchbinder, Jenny
;  TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
;  FILE REFERENCE: PA-0041 US
;  CURRENT APPLICATION NUMBER: US/09/976,594
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 60/240,409
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 1143
;  SOFTWARE: PERL Program
; SEQ ID NO 794
;   LENGTH: 1477
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6673549 4436028CB1
US-09-976-594-794

Alignment Scores:
Pred. No.:              1.09e-10        Length:         1477
Score:                  246.50          Matches:        99
Percent Similarity:     39.53%          Conservative:   69
Best Local Similarity:  23.29%          Mismatches:     172
Query Match:            5.13%           Indels:         85
DB:                     4               Gaps:           21

US-09-920-705-3 (1-901) x US-09-976-594-794 (1-1477)

Qy         4 SerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeuTrpValGly 23
Db        74 TCCGTCGCACCCACCATGATGGAAGACGACGGG---CAGCCCCGGACTCTATACGTAGGT 130

Qy        24 SerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArgTyrGly--- 42
Db       131 AACCTTTCCAGAGATGTGACAGAAGTCCTTATACTTCAGTTGTTCAGTCAGATTGGACCC 190

Qy        43 ------AspIleAspArgIleThrValTyrSerSerArgGly---PheAlaPheIle--- 58
Db       191 TGTAAAAGCTGTAAAATGATAACAGAGCATACAAGCAATGACCCATATTGCTTTGTGGAA 250

Qy        59 TyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeu 78
Db       251 TTTTATGAACAC---AGAGATGCAGCTGCTGCATTAGCTGCTATGAATGGGAGAAAAATT 307

Qy        79 AsnGlySerGlnIleLysIleGluTyrAla---ArgProAlaLysProCysLysSer--- 96
Db       308 TTGGGAAAGGAGGTCAAAGTAAACTGGGCAACCACACCAAGTAGCCAGAAAAAAGATACT 367

Qy        97 --------------LeuTrpValGlyGlyIleGlyProAsnValSerLysAspAspLeu 111
Db       368 TCCAATCACTTCCATGTGTTTGTTGGGGATTTGAGTCCAGAAATTACAACAGAAGATATC 427

Qy       112 GluGluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLys 131
Db       428 AAATCAGCATTTGCCCCCTTTGGTAAAATATCGGATGCCCGGGTAGTTAAAGACATGGCA 487

Qy       132 Thr------------------AlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
Db       488 ACTGGAAAATCCAAAGGCTATGGTTTTGTATCTTTTTATAACAAACTGGATGCAGAAAAT 547

Qy       146 Ala---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeu 164
Db       548 GCGATTGTGCATATGGGCGGTCAGTGGTTGGGTGGTCGTCAAATCCGAACCAATTGGGCC 607
```

```
Qy        165 ------ArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsn 182
Db        608 ACTCGTAAACCACCTGCACCTAAAAGTACACAA--------------GAAAACAACACT 652

Qy        183 GlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
Db        653 AAGCAGTTGAGA-----------------------TTTGAAGATGTAGTAAACCAGTCA 688

Qy        203 GlnProSerLys------ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
Db        689 AGTCCAAAAAATTGTACTGTGTACTGTGGAGGAATTGCGTCTGGGTTAACA--------- 739

Qy        221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
Db        740 ---GATCAGCTTATGAGACAGACATTCTCACCATTTGGACAAATTATGGAAATAAGAGTT 796

Qy        241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
Db        797 TTCCCAGAAAAGGGCTATTCATTTGTCAGATTTTCAACCCATGAAAGTGCAGCCCATGCC 856

Qy        261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
Db        857 ATTGTTTCGGTGAACGGTACTACGATTGAAGGACATGTGGTTAAA---TGCTATTGGGGT 913

Qy        281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
Db        914 AAAGAATCTCCT---------GATATGACTAAAAACTTCCAACAGGTTGACTATAGTCAA 964

Qy        301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
Db        965 TGG------------------------------------------------------- 967

Qy        321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsn 340
Db        968 GGCCAATGGAGCCAAGTGTATGGA-----------------AACCCACAACAGTATGGA 1009

Qy        341 AspValValGlyLysGluProAsnTrpArgArgPro----------------------- 352
Db       1010 CAGTATATGGCAAAT------GGGTGGCAAGTACCGCCTTATGGAGTATACGGGCAACCA 1063

Qy        353 ---SerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro---------GlyIle 368
Db       1064 TGGAATCAACAAGGATTTGGAGTAGATCAATCACCTTCTGCTGCTTGGATGGGTGGATTT 1123

Qy        369 LeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrpGlu 388
Db       1124 GGTGCTCAGCCTCCCCAAGGACAAGCTCCTCCCCCTGTAATACCTCCTCCTAACCAAGCC 1183

Qy        389 GlyTyrAspProAla 393
Db       1184 GGATATGGTATGGCA 1198

RESULT 5
US-07-726-607C-1
; Sequence 1, Application US/07726607C
; Patent No. 5298407
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, Paul J.
;    APPLICANT:  Streuli, Michel
;    APPLICANT:  Schlossman, Stuart F.
;    TITLE OF INVENTION:  DNA ENCODING A PROTEIN
;    TITLE OF INVENTION:  ACTIVE IN LYMPHOCYTE-
;    TITLE OF INVENTION:  MEDIATED CYTOTOXICITY
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;      SOFTWARE:  WordPerfect (Version 5.0)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/726,607C
;      FILING DATE:  July 10, 1991
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/460,678
;      FILING DATE:  January 5, 1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fraser, Janis K.
;      REGISTRATION NUMBER:  34,819
;      REFERENCE/DOCKET NUMBER:  00530/021002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2228
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-07-726-607C-1

Alignment Scores:
Pred. No.:                2.28e-10        Length:          2228
Score:                    246.00          Matches:         85
Percent Similarity:       39.65%          Conservative:    74
Best Local Similarity:    21.20%          Mismatches:      158
Query Match:              5.12%           Indels:          84
DB:                       1               Gaps:            17

US-09-920-705-3 (1-901) x US-07-726-607C-1 (1-2228)

Qy         20 LeuTrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGly 39
Db         81 CTATACGTCGGTAACCTTTCCAGAGATGTGACAGAAGCTCTAATTCTGCAACTCTTTAGC 140

Qy         40 ArgTyrGlyAspIleAspArgIle---ThrValTyrSerSerArgGly---------Phe 55
Db        141 CAGATTGGACCTTGTAAAAACTGCAAAATGATTATGGATACAGCTGGAAATGATCCCTAT 200

Qy         56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGly 75
Db        201 TGTTTTGTGGAGTTTCATGAGCATCGTCATGCAGCTGCAGCATTAGCTGCTATGAATGGA 260

Qy         76 AlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAla---ArgProAlaLysProCys 94
Db        261 CGGAAGATAATGGGTAAGGAAGTCAAAGTGAATTGGGCAACAACCCCTAGCAGTCAAAAG 320

Qy         95 LysSer-----------------LeuTrpValGlyGlyIleGlyProAsnValSerLys 108
Db        321 AAAGATACAAGCAATCATTTCCATGTCTTTGTTGGTGATCTCAGCCCAGAAATTACAACT 380

Qy        109 AspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArg 128
Db        381 GAAGATATAAAAGCTGCTTTTGCACCATTTGGAAGAATATCAGATGCCCGAGTGGTAAAA 440

Qy        129 GluArgLysThr-----------------AlaPheIleAspTyrTyrGluMetAspAsp 142
Db        441 GACATGGCAACAGGAAAGTCTAAGGGATATGGCTTTGTCTCCTTTTTCAACAAATGGGAT 500

Qy        143 AlaLeuGlnAla---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgVal 161
Db        501 GCTGAAAACGCCATTCAACAGATGGGTGGCCAGTGGCTTGGTGGAAGACAAATCAGAACT 560

Qy        162 AspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArg 181
Db        561 AACTGGGCAACCCGAAAGCCTCCCGCTCCAAAG---------AGTACATATGAGTCAAAT 611
```

```
Qy        182 AsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAsp 201
                          :::                          ||||||::::::
Db        612 ACCAAACAGCTA----------------------TCATATGATGAGGTTGTAAATCAG 647

Qy        202 ValGlnProSerLys---ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
                  ||||||      ::::::   |||          |||
Db        648 TCTAGTCCAAGCAACTGTACTGTATACTGTGGA---------GGTGTTACTTCTGGGCTA 698

Qy        221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
                 ||||||::::::                 ||||||:::|||      ::::::
Db        699 ACAGAACAACTAATGCGTCAGACTTTTTCACCATTTGGACAAATAATGGAAATTCGAGTC 758

Qy        241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
              :::|||   :::   ::::::   |||   |||   |||   |||   |||
Db        759 TTTCCAGATAAAGGATATTCATTTGTTCGGTTCAATTCCCATGAAAGTGCAGCACATGCA 818

Qy        261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
                    :::   |||                  :::|||          :::
Db        819 ATTGTTTCTGTTAATGGTACTACCATTGAAGGTCATGTTGTGAAATGCTATTGGGGCAAA 878

Qy        281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
              :::   |||
Db        879 GAAACTCTT---------------------------------------------------- 887

Qy        301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
              ||||||   |||                 |||          :::      |||   |||
Db        888 ---GATATGATAAATCCC-----------GTGCAACAGCAGAATCAAATTGGATATCCC 932

Qy        320 ------------------------------------------------------------ 320

Db        933 CAACCTTATGGCCAGTGGGGCCAGTGGTATGGAAATGCACAACAAATTGGCCAGTATATG 992

Qy        321 ------GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGlu 338
                    |||          |||       |||          :::::::::|||
Db        993 CCTAATGGTTGGCAAGTTCCTGCATATGGAATGTATGGCCAGGCATGGAACCAGCAAGGA 1052

Qy        339 TyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGly 358
              :::|||                |||   |||       |||:::          |||
Db       1053 TTTAATCAGACACAGTCTTCTGCACCA---TGGATGGGACCAAAT---------TATGGA 1100

Qy        359 IleLeuProSerProThrGlyPro------GlyIleLeuProSerProAlaGlnGlyThr 376
              :::   |||   |||   |||          :::||||||:::             |||
Db       1101 GTGCAACCG---CCTCAAGGGCAAAATGGCAGCATGTTGCCCAATCAGCCTTCTGGGTAT 1157

Qy        377 Arg 377
              |||
Db       1158 CGA 1160
```

```
RESULT 6
US-07-843-949A-1
; Sequence 1, Application US/07843949A
; Patent No. 5340935
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, Paul J.
;    APPLICANT:  Streuli, Michel
;    APPLICANT:  Schlossman, Stuart F.
;    TITLE OF INVENTION:  DNAS ENCODING PROTEINS ACTIVE IN
;    TITLE OF INVENTION:  LYMPHOCYTE-MEDIATED CYTOTOXICITY
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;      SOFTWARE:  WordPerfect (Version 5.0)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/843,949A
;      FILING DATE:  19920219
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/460,678
;      FILING DATE:  January 5, 1990
;      APPLICATION NUMBER:  07/726,607
;      FILING DATE:  July 10, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fraser, Janis K.
;      REGISTRATION NUMBER:  34,819
;      REFERENCE/DOCKET NUMBER:  00530/021003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2228
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-07-843-949A-1

Alignment Scores:
Pred. No.:                2.28e-10        Length:          2228
Score:                    246.00          Matches:         85
Percent Similarity:       39.65%          Conservative:    74
Best Local Similarity:    21.20%          Mismatches:      158
Query Match:              5.12%           Indels:          84
DB:                       1               Gaps:            17

US-09-920-705-3 (1-901) x US-07-843-949A-1 (1-2228)

Qy         20 LeuTrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGly 39
              |||:::|||||:::|||:::   :::   ||||||:::   :::   :::||||||
Db         81 CTATACGTCGGTAACCTTTCCAGAGATGTGACAGAAGCTCTAATTCTGCAACTCTTTAGC 140

Qy         40 ArgTyrGlyAspIleAspArgIle---ThrValTyrSerSerArgGly---------Phe 55
              :::   |||                 :::   :::   |||          :::
Db        141 CAGATTGGACCTTGTAAAAACTGCAAAATGATTATGGATACAGCTGGAAATGATCCCTAT 200

Qy         56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGly 75
                 |||:::   :::                 |||   ||||||      |||:::   |||
Db        201 TGTTTTGTGGAGTTTCATGAGCATCGTCATGCAGCTGCAGCATTAGCTGCTATGAATGGA 260

Qy         76 AlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAla---ArgProAlaLysProCys 94
                    :::   |||   ::::::|||:::   :::|||      |||:::
Db        261 CGGAAGATAATGGGTAAGGAAGTCAAAGTGAATTGGGCAACAACCCCTAGCAGTCAAAAG 320

Qy         95 LysSer-----------------LeuTrpValGlyGlyIleGlyProAsnValSerLys 108
              |||                  ::::::||||||   :::   |||   ::::::
Db        321 AAAGATACAAGCAATCATTTCCATGTCTTTGTTGGTGATCTCAGCCCAGAAATTACAACT 380

Qy        109 AspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArg 128
              :::|||::::::      |||:::   ||||||:::|||   |||   |||   ::::::
Db        381 GAAGATATAAAAGCTGCTTTTGCACCATTTGGAAGAATATCAGATGCCCGAGTGGTAAAA 440

Qy        129 GluArgLysThr-----------------AlaPheIleAspTyrTyrGluMetAspAsp 142
              :::      |||                |||:::   ::::::          |||
Db        441 GACATGGCAACAGGAAAGTCTAAGGGATATGGCTTTGTCTCCTTTTTCAACAAATGGGAT 500

Qy        143 AlaLeuGlnAla---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgVal 161
              |||      |||   :::   |||   |||:::   :::||||||      :::|||
Db        501 GCTGAAAACGCCATTCAACAGATGGGTGGCCAGTGGCTTGGTGGAAGACAAATCAGAACT 560

Qy        162 AspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArg 181
              ::::::          :::   |||       :::            :::|||::::::
Db        561 AACTGGGCAACCCGAAAGCCTCCCGCTCCAAAG---------AGTACATATGAGTCAAAT 611

Qy        182 AsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAsp 201
                          :::                          ||||||::::::
Db        612 ACCAAACAGCTA----------------------TCATATGATGAGGTTGTAAATCAG 647
```

```
Qy        202 ValGlnProSerLys---ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
Db        648 TCTAGTCCAAGCAACTGTACTGTATACTGTGGA---------GGTGTTACTTCTGGGCTA 698

Qy        221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
Db        699 ACAGAACAACTAATGCGTCAGACTTTTTCACCATTTGGACAAATAATGGAAATTCGAGTC 758

Qy        241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
Db        759 TTTCCAGATAAAGGATATTCATTTGTTCGGTTCAATTCCCATGAAAGTGCAGCACATGCA 818

Qy        261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
Db        819 ATTGTTTCTGTTAATGGTACTACCATTGAAGGTCATGTTGTGAAATGCTATTGGGGCAAA 878

Qy        281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
Db        879 GAAACTCTT---------------------------------------------------- 887

Qy        301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
Db        888 ---GATATGATAAATCCC------------GTGCAACAGCAGAATCAAATTGGATATCCC 932

Qy        320 ------------------------------------------------------------ 320
Db        933 CAACCTTATGGCCAGTGGGGCCAGTGGTATGGAAATGCACAACAAATTGGCCAGTATATG 992

Qy        321 ------GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGlu 338
Db        993 CCTAATGGTTGGCAAGTTCCTGCATATGGAATGTATGGCCAGGCATGGAACCAGCAAGGA 1052

Qy        339 TyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGly 358
Db       1053 TTTAATCAGACACAGTCTTCTGCACCA---TGGATGGGACCAAAT---------TATGGA 1100

Qy        359 IleLeuProSerProThrGlyPro------GlyIleLeuProSerProAlaGlnGlyThr 376
Db       1101 GTGCAACCG---CCTCAAGGGCAAAATGGCAGCATGTTGCCCAATCAGCCTTCTGGGTAT 1157

Qy        377 Arg 377
Db       1158 CGA 1160
```

```
RESULT 7
US-08-218-978-1
; Sequence 1, Application US/08218978
; Patent No. 5837811
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, Paul J.
;    APPLICANT:  Streuli, Michel
;    APPLICANT:  Schlossman, Stuart F.
;    TITLE OF INVENTION:  DNAS ENCODING PROTEINS ACTIVE IN
;    TITLE OF INVENTION:  LYMPHOCYTE-MEDIATED CYTOTOXICITY
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;      SOFTWARE:  WordPerfect (Version 5.0)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/218,978
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/843,949
;      FILING DATE:  February 19, 1992
;      APPLICATION NUMBER:  07/460,678
;      FILING DATE:  January 5, 1990
;      APPLICATION NUMBER:  07/726,607
;      FILING DATE:  July 10, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fraser, Janis K.
;      REGISTRATION NUMBER:  34,819
;      REFERENCE/DOCKET NUMBER:  00530/021003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2228
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-218-978-1

Alignment Scores:
Pred. No.:                2.28e-10      Length:          2228
Score:                    246.00        Matches:         85
Percent Similarity:       39.65%        Conservative:    74
Best Local Similarity:    21.20%        Mismatches:      158
Query Match:              5.12%         Indels:          84
DB:                       2             Gaps:            17

US-09-920-705-3 (1-901) x US-08-218-978-1 (1-2228)

Qy         20 LeuTrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGly 39
Db         81 CTATACGTCGGTAACCTTTCCAGAGATGTGACAGAAGCTCTAATTCTGCAACTCTTTAGC 140

Qy         40 ArgTyrGlyAspIleAspArgIle---ThrValTyrSerSerArgGly---------Phe 55
Db        141 CAGATTGGACCTTGTAAAAACTGCAAAATGATTATGGATACAGCTGGAAATGATCCCTAT 200

Qy         56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGly 75
Db        201 TGTTTTGTGGAGTTTCATGAGCATCGTCATGCAGCTGCAGCATTAGCTGCTATGAATGGA 260

Qy         76 AlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAla---ArgProAlaLysProCys 94
Db        261 CGGAAGATAATGGGTAAGGAAGTCAAAGTGAATTGGGCAACAACCCCTAGCAGTCAAAAG 320

Qy         95 LysSer------------------LeuTrpValGlyGlyIleGlyProAsnValSerLys 108
Db        321 AAAGATACAAGCAATCATTTCCATGTCTTTGTTGGTGATCTCAGCCCAGAAATTACAACT 380

Qy        109 AspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArg 128
Db        381 GAAGATATAAAAGCTGCTTTTGCACCATTTGGAAGAATATCAGATGCCCGAGTGGTAAAA 440

Qy        129 GluArgLysThr------------------AlaPheIleAspTyrTyrGluMetAspAsp 142
Db        441 GACATGGCAACAGGAAAGTCTAAGGGATATGGCTTTGTCTCCTTTTTCAACAAATGGGAT 500

Qy        143 AlaLeuGlnAla---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgVal 161
Db        501 GCTGAAAACGCCATTCAACAGATGGGTGGCCAGTGGCTTGGTGGAAGACAAATCAGAACT 560

Qy        162 AspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArg 181
Db        561 AACTGGGCAACCCGAAAGCCTCCCGCTCCAAAG---------AGTACATATGAGTCAAAT 611

Qy        182 AsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAsp 201
Db        612 ACCAAACAGCTA----------------------TCATATGATGAGGTTGTAAATCAG 647

Qy        202 ValGlnProSerLys---ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
```

```
Db        648 TCTAGTCCAAGCAACTGTACTGTATACTGTGGA---------GGTGTTACTTCTGGGCTA 698

Qy        221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240

Db        699 ACAGAACAACTAATGCGTCAGACTTTTTCACCATTTGGACAAATAATGGAAATTCGAGTC 758

Qy        241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260

Db        759 TTTCCAGATAAAGGATATTCATTTGTTCGGTTCAATTCCCATGAAAGTGCAGCACATGCA 818

Qy        261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280

Db        819 ATTGTTTCTGTTAATGGTACTACCATTGAAGGTCATGTTGTGAAATGCTATTGGGGCAAA 878

Qy        281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300

Db        879 GAAACTCTT---------------------------------------------------- 887

Qy        301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320

Db        888 ---GATATGATAAATCCC------------GTGCAACAGCAGAATCAAATTGGATATCCC 932

Qy        320 ------------------------------------------------------------ 320

Db        933 CAACCTTATGGCCAGTGGGGCCAGTGGTATGGAAATGCACAACAAATTGGCCAGTATATG 992

Qy        321 ------GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGlu 338

Db        993 CCTAATGGTTGGCAAGTTCCTGCATATGGAATGTATGGCCAGGCATGGAACCAGCAAGGA 1052

Qy        339 TyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGly 358

Db       1053 TTTAATCAGACACAGTCTTCTGCACCA---TGGATGGGACCAAAT---------TATGGA 1100

Qy        359 IleLeuProSerProThrGlyPro------GlyIleLeuProSerProAlaGlnGlyThr 376

Db       1101 GTGCAACCG---CCTCAAGGGCAAAATGGCAGCATGTTGCCCAATCAGCCTTCTGGGTAT 1157

Qy        377 Arg 377

Db       1158 CGA 1160
```

RESULT 8
US-09-949-016-4865
; Sequence 4865, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4865
; LENGTH: 2356
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4865

Alignment Scores:

Pred. No.:	2.48e-10	Length:	2356
Score:	246.00	Matches:	85
Percent Similarity:	39.65%	Conservative:	74
Best Local Similarity:	21.20%	Mismatches:	158
Query Match:	5.12%	Indels:	84

```
DB:                    4              Gaps:          17

US-09-920-705-3 (1-901) x US-09-949-016-4865 (1-2356)

Qy         20 LeuTrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGly 39

Db        210 CTATACGTCGGTAACCTTTCCAGAGATGTGACAGAAGCTCTAATTCTGCAACTCTTTAGC 269

Qy         40 ArgTyrGlyAspIleAspArgIle---ThrValTyrSerSerArgGly---------Phe 55

Db        270 CAGATTGGACCTTGTAAAAACTGCAAAATGATTATGGATACAGCTGGAAATGATCCCTAT 329

Qy         56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGly 75

Db        330 TGTTTTGTGGAGTTTCATGAGCATCGTCATGCAGCTGCAGCATTAGCTGCTATGAATGGA 389

Qy         76 AlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAla---ArgProAlaLysProCys 94

Db        390 CGGAAGATAATGGGTAAGGAAGTCAAAGTGAATTGGGCAACAACCCCTAGCAGTCAAAAG 449

Qy         95 LysSer-----------------LeuTrpValGlyGlyIleGlyProAsnValSerLys 108

Db        450 AAAGATACAAGCAATCATTTCCATGTCTTTGTTGGTGATCTCAGCCCAGAAATTACAACT 509

Qy        109 AspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArg 128

Db        510 GAAGATATAAAAGCTGCTTTTGCACCATTTGGAAGAATATCAGATGCCCGAGTGGTAAAA 569

Qy        129 GluArgLysThr-----------------AlaPheIleAspTyrTyrGluMetAspAsp 142

Db        570 GACATGGCAACAGGAAAGTCTAAGGGATATGGCTTTGTCTCCTTTTTCAACAAATGGGAT 629

Qy        143 AlaLeuGlnAla---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgVal 161

Db        630 GCTGAAAACGCCATTCAACAGATGGGTGGCCAGTGGCTTGGTGGAAGACAAATCAGAACT 689

Qy        162 AspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArg 181

Db        690 AACTGGGCAACCCGAAAGCCTCCCGCTCCAAAG---------AGTACATATGAGTCAAAT 740

Qy        182 AsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAsp 201

Db        741 ACCAAACAGCTA-----------------------TCATATGATGAGGTTGTAAATCAG 776

Qy        202 ValGlnProSerLys---ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220

Db        777 TCTAGTCCAAGCAACTGTACTGTATACTGTGGA---------GGTGTTACTTCTGGGCTA 827

Qy        221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240

Db        828 ACAGAACAACTAATGCGTCAGACTTTTTCACCATTTGGACAAATAATGGAAATTCGAGTC 887

Qy        241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260

Db        888 TTTCCAGATAAAGGATATTCATTTGTTCGGTTCAATTCCCATGAAAGTGCAGCACATGCA 947

Qy        261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280

Db        948 ATTGTTTCTGTTAATGGTACTACCATTGAAGGTCATGTTGTGAAATGCTATTGGGGCAAA 1007

Qy        281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300

Db       1008 GAAACTCTT---------------------------------------------------- 1016

Qy        301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320

Db       1017 ---GATATGATAAATCCC------------GTGCAACAGCAGAATCAAATTGGATATCCC 1061

Qy        320 ------------------------------------------------------------ 320

Db       1062 CAACCTTATGGCCAGTGGGGCCAGTGGTATGGAAATGCACAACAAATTGGCCAGTATATG 1121

Qy        321 ------GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGlu 338
```

```
Db       1122 CCTAATGGTTGGCAAGTTCCTGCATATGGAATGTATGGCCAGGCATGGAACCAGCAAGGA 1181

Qy        339 TyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGly 358
Db       1182 TTTAATCAGACACAGTCTTCTGCACCA---TGGATGGGACCAAAT---------TATGGA 1229

Qy        359 IleLeuProSerProThrGlyPro------GlyIleLeuProSerProAlaGlnGlyThr 376
Db       1230 GTGCAACCG---CCTCAAGGGCAAAATGGCAGCATGTTGCCCAATCAGCCTTCTGGGTAT 1286

Qy        377 Arg 377
Db       1287 CGA 1289
```

RESULT 9
US-09-949-016-4866
; Sequence 4866, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4866
; LENGTH: 2356
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4866

Alignment Scores:

Pred. No.:	2.48e-10	Length:	2356
Score:	246.00	Matches:	85
Percent Similarity:	39.65%	Conservative:	74
Best Local Similarity:	21.20%	Mismatches:	158
Query Match:	5.12%	Indels:	84
DB:	4	Gaps:	17

US-09-920-705-3 (1-901) x US-09-949-016-4866 (1-2356)

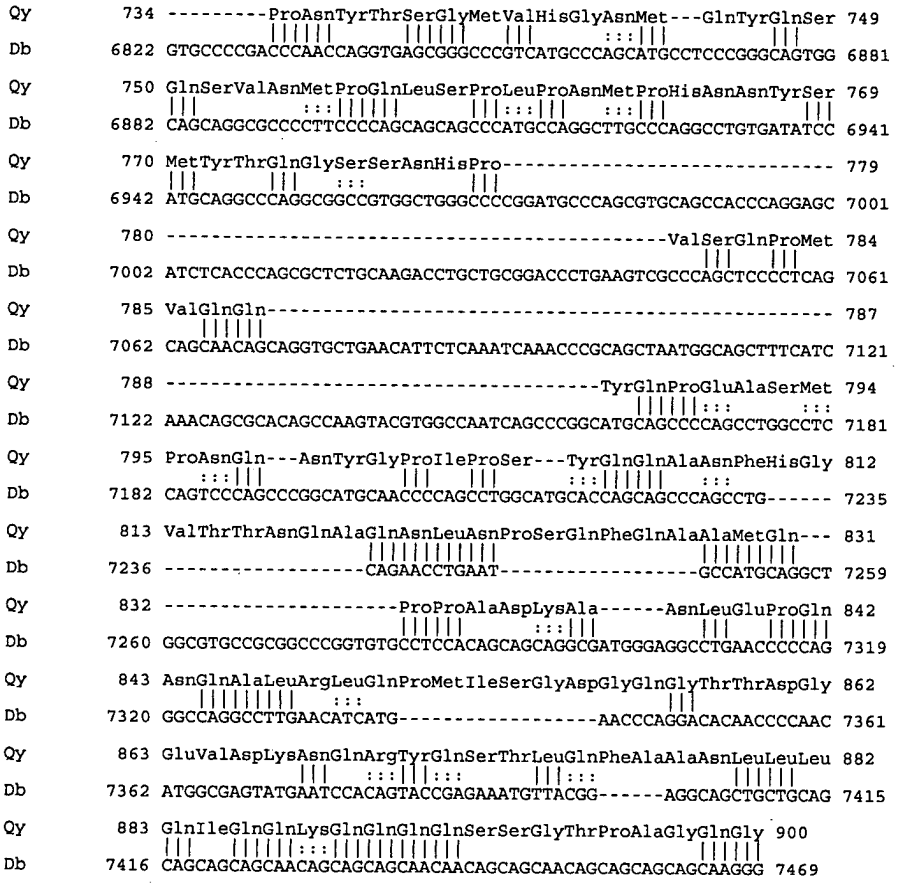
```
Qy         20 LeuTrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGly 39
Db        210 CTATACGTCGGTAACCTTTCCAGAGATGTGACAGAAGCTCTAATTCTGCAACTCTTTAGC 269

Qy         40 ArgTyrGlyAspIleAspArgIle---ThrValTyrSerSerArgGly---------Phe 55
Db        270 CAGATTGGACCTTGTAAAAACTGCAAAATGATTATGGATACAGCTGGAAATGATCCCTAT 329

Qy         56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGly 75
Db        330 TGTTTTGTGGAGTTTCATGAGCATCGTCATGCAGCTGCAGCATTAGCTGCTATGAATGGA 389

Qy         76 AlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAla---ArgProAlaLysProCys 94
Db        390 CGGAAGATAATGGGTAAGGAAGTCAAAGTGAATTGGGCAACAACCCCTAGCAGTCAAAAG 449

Qy         95 LysSer-----------------LeuTrpValGlyGlyIleGlyProAsnValSerLys 108
Db        450 AAAGATACAAGCAATCATTTCCATGTCTTTGTTGGTGATCTCAGCCCAGAAATTACAACT 509

Qy        109 AspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArg 128
Db        510 GAAGATATAAAAGCTGCTTTTGCACCATTTGGAAGAATATCAGATGCCCGAGTGGTAAAA 569

Qy        129 GluArgLysThr-----------------AlaPheIleAspTyrTyrGluMetAspAsp 142
Db        570 GACATGGCAACAGGAAAGTCTAAGGGATATGGCTTTGTCTCCTTTTTCAACAAATGGGAT 629

Qy        143 AlaLeuGlnAla---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgVal 161
Db        630 GCTGAAAACGCCATTCAACAGATGGGTGGCCAGTGGCTTGGTGGAAGACAAATCAGAACT 689

Qy        162 AspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArg 181
Db        690 AACTGGGCAACCCGAAAGCCTCCCGCTCCAAAG---------AGTACATATGAGTCAAAT 740

Qy        182 AsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAsp 201
Db        741 ACCAAACAGCTA----------------------TCATATGATGAGGTTGTAAATCAG 776

Qy        202 ValGlnProSerLys---ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
Db        777 TCTAGTCCAAGCAACTGTACTGTATACTGTGGA---------GGTGTTACTTCTGGGCTA 827

Qy        221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
Db        828 ACAGAACAACTAATGCGTCAGACTTTTTCACCATTTGGACAAATAATGGAAATTCGAGTC 887

Qy        241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
Db        888 TTTCCAGATAAAGGATATTCATTTGTTCGGTTCAATTCCCATGAAAGTGCAGCACATGCA 947

Qy        261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
Db        948 ATTGTTTCTGTTAATGGTACTACCATTGAAGGTCATGTTGTGAAATGCTATTGGGGCAAA 1007

Qy        281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
Db       1008 GAAACTCTT---------------------------------------------------- 1016

Qy        301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
Db       1017 ---GATATGATAAATCCC-----------GTGCAACAGCAGAATCAAATTGGATATCCC 1061

Qy        320 ------------------------------------------------------------ 320
Db       1062 CAACCTTATGGCCAGTGGGGCCAGTGGTATGGAAATGCACAACAAATTGGCCAGTATATG 1121

Qy        321 ------GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGlu 338
Db       1122 CCTAATGGTTGGCAAGTTCCTGCATATGGAATGTATGGCCAGGCATGGAACCAGCAAGGA 1181

Qy        339 TyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGly 358
Db       1182 TTTAATCAGACACAGTCTTCTGCACCA---TGGATGGGACCAAAT---------TATGGA 1229

Qy        359 IleLeuProSerProThrGlyPro------GlyIleLeuProSerProAlaGlnGlyThr 376
Db       1230 GTGCAACCG---CCTCAAGGGCAAAATGGCAGCATGTTGCCCAATCAGCCTTCTGGGTAT 1286

Qy        377 Arg 377
Db       1287 CGA 1289
```

RESULT 10
US-07-843-949A-3
; Sequence 3, Application US/07843949A
; Patent No. 5340935
; GENERAL INFORMATION:
; APPLICANT: Anderson, Paul J.
; APPLICANT: Streuli, Michel
; APPLICANT: Schlossman, Stuart F.
; TITLE OF INVENTION: DNAS ENCODING PROTEINS ACTIVE IN
; TITLE OF INVENTION: LYMPHOCYTE-MEDIATED CYTOTOXICITY
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:

```
;       ADDRESSEE:  Fish & Richardson
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       COMPUTER:  IBM PS/2 Model 50Z or 55SX
;       OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;       SOFTWARE:  WordPerfect (Version 5.0)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/843,949A
;       FILING DATE:  19920219
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/460,678
;       FILING DATE:  January 5, 1990
;       APPLICATION NUMBER:  07/726,607
;       FILING DATE:  July 10, 1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fraser, Janis K.
;       REGISTRATION NUMBER:  34,819
;       REFERENCE/DOCKET NUMBER:  00530/021003
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 542-5070
;       TELEFAX:  (617) 542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1401
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-07-843-949A-3

Alignment Scores:
Pred. No.:                2.14e-10        Length:          1401
Score:                    242.50          Matches:         99
Percent Similarity:       39.29%          Conservative:    68
Best Local Similarity:    23.29%          Mismatches:      173
Query Match:              5.05%           Indels:          85
DB:                       1               Gaps:            21

US-09-920-705-3 (1-901) x US-07-843-949A-3 (1-1401)

Qy          4 SerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeuTrpValGly 23
              |||:::   |||         :::|||   |||   |||         |||:::||||||
Db         31 TCCGTCGCACCCACCATGATGGAAGACGACGGG---CAGCCCCGGACTCTATACGTAGGT 87

Qy         24 SerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArgTyrGly--- 42
              :::|||:::   :::   ||||||      :::   :::||||||   :::   |||
Db         88 AACCTTTCCAGAGATGTGACAGAAGTCCTTATACTTCAGTTGTTCAGTCAGATTGGACCC 147

Qy         43 ------AspIleAspArgIleThrValTyrSerSerArgGly---PheAlaPheIle--- 58
                     ||||||   ::::::|||        :::   |||:::
Db        148 TGTAAAAGCTGTAAAATGATAACAGAGCATACAAGCAATGACCCATATTGCTTTGTGGAA 207

Qy         59 TyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeu 78
              :::|||   |||       :::|||   ||||||      |||:::   |||       :::
Db        208 TTTTATGAACAC---AGAGATGCAGCTGCTGCATTAGCTGCTATGAATGGGAGAAAAATT 264

Qy         79 AsnGlySerGlnIleLysIleGluTyrAla---ArgProAlaLysProCysLysSer--- 96
                 |||   ::::::|||:::   :::|||       |||:::         |||
Db        265 TTGGGAAAGGAGGTCAAAGTAAACTGGGCAACCACACCAAGTAGCCAGAAAAAAGATACT 324

Qy         97 --------------LeuTrpValGlyGlyIleGlyProAsnValSerLysAspAspLeu 111
                           ::::::||||||   :::   |||   ::::::   :::|||:::
Db        325 TCCAATCACTTCCATGTGTTTGTTGGGGATTTGAGTCCAGAAATTACAACAGAAGATATC 384

Qy        112 GluGluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLys 131
              :::       |||:::   ||||||||||||   |||   |||   :::::::::
Db        385 AAATCAGCATTTGCCCCCTTTGGTAAAATATCGGATGCCCGGGTAGTTAAAGACATGGCA 444

Qy        132 Thr------------------AlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
              |||                     |||:::   :::|||         ||||||
Db        445 ACTGGAAAATCCAAAGGCTATGGTTTTGTATCTTTTTATAACAAACTGGATGCAGAAAAT 504

Qy        146 Ala---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeu 164
              |||         |||   |||:::   :::||||||       :::|||   ::::::
Db        505 GCGATTGTGCATATGGGCGGTCAGTGGTTGGGTGGTCGTCAAATCCGAACCAATTGGGCC 564

Qy        165 ------ArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsn 182
                    :::       |||||||||       |||                 :::|||
Db        565 ACTCGTAAACCACCTGCACCTAAAAGTACACAA--------------GAAAACAACACT 609

Qy        183 GlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
                    :::                             :::||||||
Db        610 AAGCAGTTGAGA----------------------TTTGAAGATGTAGTAAACCAGTCA 645

Qy        203 GlnProSerLys------ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
                 |||            |||         |||               |||
Db        646 AGTCCAAAAAATTGTACTGTGTACTGTGGAGGAATTGCGTCTGGGTTAACA--------- 696

Qy        221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
                 :::|||::::::                    ||||||:::|||      ::::::
Db        697 ---GATCAGCTTATGAGACAGACATTCTCACCATTTGGACAAATTATGGAAATAAGAGTT 753

Qy        241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
                 |||   :::   ::::::   |||   |||   :::   |||   |||
Db        754 TTGCCAGAAAAGGGCTATTCATTTGTCAGATTTTCAACCCATGAAAGTGCAGCCCATGCC 813

Qy        261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
                    :::   |||                       :::|||       |||
Db        814 ATTGTTTCGGTGAACGGTACTACGATTGAAGGACATGTGGTTAAA---TGCTATTGGGGT 870

Qy        281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
                 |||   |||         |||   |||       :::       :::       |||:::
Db        871 AAAGAATCTCCT---------GATATGACTAAAAACTTCCAACAGGTTGACTATAGTCAA 921

Qy        301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320

Db        922 TGG------------------------------------------------------- 924

Qy        321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsn 340
              |||             :::   |||                   |||       :::|||
Db        925 GGCCAATGGAGCCAAGTGTATGGA-----------------AACCCACAACAGTATGGA 966

Qy        341 AspValValGlyLysGluProAsnTrpArgArgPro----------------------- 352
                          :::                 |||:::   |||
Db        967 CAGTATATGGCAAAT------GGGTGGCAAGTACCGCCTTATGGAGTATACGGGCAACCA 1020

Qy        353 ---SerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro---------GlyIle 368
                 :::       |||   |||:::       |||||||:::                |||
Db       1021 TGGAATCAACAAGGATTTGGAGTAGATCAATCACCTTCTGCTGCTTGGATGGGTGGATTT 1080

Qy        369 LeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrpGlu 388
                    |||   ||||||           |||               |||   :::
Db       1081 GGTGCTCAGCCTCCCCAAGGACAAGCTCCTCCCCCTGTAATACCTCCTCCTAACCAAGCC 1140

Qy        389 GlyTyrAspProAla 393
              ||||||      |||
Db       1141 GGATATGGTATGGCA 1155

RESULT 11
US-08-218-978-3
; Sequence 3, Application US/08218978
; Patent No. 5837811
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, Paul J.
;    APPLICANT:  Streuli, Michel
;    APPLICANT:  Schlossman, Stuart F.
;    TITLE OF INVENTION:  DNAS ENCODING PROTEINS ACTIVE IN
;    TITLE OF INVENTION:  LYMPHOCYTE-MEDIATED CYTOTOXICITY
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  Fish & Richardson
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       COMPUTER:  IBM PS/2 Model 50Z or 55SX
;       OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;       SOFTWARE:  WordPerfect (Version 5.0)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/218,978
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/843,949
;       FILING DATE:  February 19, 1992
;       APPLICATION NUMBER:  07/460,678
;       FILING DATE:  January 5, 1990
;       APPLICATION NUMBER:  07/726,607
;       FILING DATE:  July 10, 1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fraser, Janis K.
;       REGISTRATION NUMBER:  34,819
;       REFERENCE/DOCKET NUMBER:  00530/021003
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 542-5070
;       TELEFAX:  (617) 542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1401
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-218-978-3

Alignment Scores:
Pred. No.:                2.14e-10        Length:          1401
Score:                    242.50          Matches:         99
Percent Similarity:       39.29%          Conservative:    68
Best Local Similarity:    23.29%          Mismatches:      173
Query Match:              5.05%           Indels:          85
DB:                       2               Gaps:            21

US-09-920-705-3 (1-901) x US-08-218-978-3 (1-1401)

Qy          4 SerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeuTrpValGly 23
              ||:::   |||         :::|||   |||   |||         |||:::||||||
Db         31 TCCGTCGCACCCACCATGATGGAAGACGACGGG---CAGCCCCGGACTCTATACGTAGGT 87

Qy         24 SerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArgTyrGly--- 42
              :::|||:::   :::  ||||||    :::   :::||||||   :::   |||
Db         88 AACCTTTCCAGAGATGTGACAGAAGTCCTTATACTTCAGTTGTTCAGTCAGATTGGACCC 147

Qy         43 ------AspIleAspArgIleThrValTyrSerSerArgGly---PheAlaPheIle--- 58
                    ||||||   ::::::|||         :::   |||:::
Db        148 TGTAAAAGCTGTAAAATGATAACAGAGCATACAAGCAATGACCCATATTGCTTTGTGGAA 207

Qy         59 TyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeu 78
              :::|||   |||      :::|||   ||||||      |||:::   |||      :::
Db        208 TTTTATGAACAC---AGAGATGCAGCTGCTGCATTAGCTGCTATGAATGGGAGAAAAATT 264

Qy         79 AsnGlySerGlnIleLysIleGluTyrAla---ArgProAlaLysProCysLysSer--- 96
                 |||   ::::::|||:::   :::|||      |||:::        |||
Db        265 TTGGGAAAGGAGGTCAAAGTAAACTGGGCAACCACACCAAGTAGCCAGAAAAAAGATACT 324

Qy         97 --------------LeuTrpValGlyGlyIleGlyProAsnValSerLysAspAspLeu 111
                           ::::::||||||   :::   |||   ::::::   :::|||:::
Db        325 TCCAATCACTTCCATGTGTTTGTTGGGGATTTGAGTCCAGAAATTACAACAGAAGATATC 384

Qy        112 GluGluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLys 131
              :::      |||:::   ||||||||||||   |||   |||   :::::::::
Db        385 AAATCAGCATTTGCCCCCTTTGGTAAAATATCGGATGCCCGGGTAGTTAAAGACATGGCA 444

Qy        132 Thr------------------AlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
              |||                 |||:::   :::|||           ||||||
Db        445 ACTGGAAAATCCAAAGGCTATGGTTTTGTATCTTTTTATAACAAACTGGATGCAGAAAAT 504

Qy        146 Ala---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeu 164
              |||      |||   |||:::   :::|||||||       :::|||   ::::::
Db        505 GCGATTGTGCATATGGGCGGTCAGTGGTTGGGTGGTCGTCAAATCCGAACCAATTGGGCC 564

Qy        165 ------ArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsn 182
                    :::      |||||||||       |||                 :::|||
Db        565 ACTCGTAAACCACCTGCACCTAAAAGTACACAA--------------GAAAACAACACT 609

Qy        183 GlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
                    :::                         :::||||||
Db        610 AAGCAGTTGAGA----------------------TTTGAAGATGTAGTAAACCAGTCA 645

Qy        203 GlnProSerLys------ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
                 |||            |||      |||             |||
Db        646 AGTCCAAAAAATTGTACTGTGTACTGTGGAGGAATTGCGTCTGGGTTAACA--------- 696

Qy        221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
                 :::|||::::::                  ||||||:::|||      ::::::
Db        697 ---GATCAGCTTATGAGACAGACATTCTCACCATTTGGACAAATTATGGAAATAAGAGTT 753

Qy        241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
                 |||   :::   ::::::   |||   |||   :::   |||   |||
Db        754 TTGCCAGAAAAGGGCTATTCATTTGTCAGATTTTCAACCCATGAAAGTGCAGCCCATGCC 813

Qy        261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
                    :::   |||                     :::|||      |||
Db        814 ATTGTTTCGGTGAACGGTACTACGATTGAAGGACATGTGGTTAAA---TGCTATTGGGGT 870

Qy        281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
                 |||   |||         |||   |||      :::      :::.      |||:::
Db        871 AAAGAATCTCCT---------GATATGACTAAAAACTTCCAACAGGTTGACTATAGTCAA 921

Qy        301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320

Db        922 TGG---------------------------------------------------------- 924

Qy        321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsn 340
              |||             :::   |||                |||      :::|||
Db        925 GGCCAATGGAGCCAAGTGTATGGA-----------------AACCCACAACAGTATGGA 966

Qy        341 AspValValGlyLysGluProAsnTrpArgArgPro----------------------- 352
                    :::                |||::   |||
Db        967 CAGTATATGGCAAAT------GGGTGGCAAGTACCGCCTTATGGAGTATACGGGCAACCA 1020

Qy        353 ---SerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro---------GlyIle 368
                 :::      |||   |||:::      ||||||:::              |||
Db       1021 TGGAATCAACAAGGATTTGGAGTAGATCAATCACCTTCTGCTGCTTGGATGGGTGGATTT 1080

Qy        369 LeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrpGlu 388
                   |||   ||||||          |||             |||   :::
Db       1081 GGTGCTCAGCCTCCCCAAGGACAAGCTCCTCCCCCTGTAATACCTCCTCCTAACCAAGCC 1140

Qy        389 GlyTyrAspProAla 393
              ||||||      |||
Db       1141 GGATATGGTATGGCA 1155

RESULT 12
US-09-514-247A-9
; Sequence 9, Application US/09514247A
; Patent No. 6365361
; GENERAL INFORMATION:
;  APPLICANT: TANABE SEIYAKU CO. LTD.
;  APPLICANT:  TANIGUCHI, Tomoyasu
;  APPLICANT:  MIZUKAMI, Junko
;  TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND ANTAGONIST TO PPA
;  FILE REFERENCE: TANIGUCHI=6
```

```
;  CURRENT APPLICATION NUMBER: US/09/514,247A
;  CURRENT FILING DATE:  2000-02-28
;  PRIOR APPLICATION NUMBER: PCT/JP98/03734
;  PRIOR FILING DATE: 1998-08-24
;  PRIOR APPLICATION NUMBER: JP231084/1997
;  PRIOR FILING DATE: 1997-08-27
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
;   LENGTH: 8147
;   TYPE: DNA
;   ORGANISM: human
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (819)..(8147)
US-09-514-247A-9

Alignment Scores:
Pred. No.:              1.06e-08      Length:         8147
Score:                  236.50        Matches:        203
Percent Similarity:     32.86%        Conservative:   73
Best Local Similarity:  24.17%        Mismatches:     297
Query Match:            4.93%         Indels:         269
DB:                     3             Gaps:           43

US-09-920-705-3 (1-901) x US-09-514-247A-9 (1-8147)

Qy        251 PheArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsn 270
Db       5189 TTTAGAGTATGTGAAGAAATTAGGGTATGTGA-CAGGGCACATCTGGGCCTGTCCTC--- 5244

Qy        271 AsnProArgIleLysIleMetTyrSerAsnAsp-----------------GluLeuPro 284
Db       5245 ------CAAGTGAAGGAGATGATTACATCTTCCATTGCCACCCACCTGATCAAAAAATAC 5298

Qy        285 ProGluGlnAspAspThrSerPheTyrSerGlyMetLysArg----------------- 298
Db       5299 CCAAGCCAAAACGACTGCAGG------AGTGGTACAAAAAGATGCTGGACAAGGCGTTTG 5352

Qy        299 --------------SerArgThrAspMetPheAsnAsnAspProSerCysValSerSer 313
Db       5353 CAGAGCGGATCATCCATGACTACAAGGATATTTTCAAACAAGCAACTGAAGACAGGCTCA 5412

Qy        314 ProHisSerThrGlyIlePro-----------------GlySerMet---------Arg 324
Db       5413 CCAGTGCCAAGGAACTGCCCTATTTTGAAGGTGATTTCTGGCCCAATGTGTTAGAAGAGA 5472

Qy        325 ProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValValGly 344
Db       5473 GCATTAAGGAACTAGAACAAGAAGAAGAGGAGAGGAAAAAGGAAGAGAGCACTGCAGCCA 5532

Qy        345 LysGluPro-------------------------------------------AsnTrpArg 350
Db       5533 GTGAAACCACTGAGGGCAGTCAGGGCGACAGCAAGAATGCCAAGAAGAAGAACAACAAGA 5592

Qy        351 ArgPro-----------------SerAlaAsnGlyThrGlyIleLeuProSer---Pro 363
Db       5593 AAACCAACAAGAACAAAAGCAGCATCAGCCGCGCCAACAAGAAGAAGCCCAGCATGCCCA 5652

Qy        364 ThrGlyProGlyIleLeuPro-----------SerProAlaGlnGlyThrArgArg--- 378
Db       5653 ACGTGTCCAATGACCTGTCCCAGAAGCTGTATGCCACCATGGAGAAGCACAAGGAGGTCT 5712

Qy        379 ---------------------------------------------ProMetArgSerAsn 383
Db       5713 TCTTCGTGATCCACCTGCACGCTGGGCCTGTCATCAACACCCTGCCCCCCATCGTCGACC 5772

Qy        384 Pro-----------------AspSerTrpGluGlyTyrAspProAlaGlnLeuValArg 397
Db       5773 CCGACCCCCTGCTCAGCTGTGACCTCATGGATGGGCGCGACGCCTTCCTCA--------- 5823

Qy        398 GluSerLysArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAsp 417

Db       5824 CCCTCGCCAGAGACAAGCACTGGGAGTTCTCCTCCTTGCGCCGCTCCAAGTGGTCCACGC 5883

Qy        418 GluArgSerPheGlyArgGlySerValAlaAlaArgProIleArgGlyProPro------ 435
Db       5884 TCTGCATGCTGGTGGAGCTGCACACCCAGGGCCAGGACCGCTTTGTCTACACCTGCAACG 5943

Qy        436 AspSerAspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrProValCysCys--- 454
Db       5944 AGTGCAAGCACCACGTGGAGA-----------CGCGCTGGCACTGCACTGTGTGCGAGG 5991

Qy        455 ----------------------------AlaArgCysValProMet------------ 460
Db       5992 ACTACGACCTCTGCATCAACTGCTATAACACGAAGAGCCATGCCCATAAGATGGTGAAGT 6051

Qy        461 ---GlyLysGlyIleGluThrLysLeuProGluValValAsnCysSerAlaArgThrAsp 479
Db       6052 GGGGGCTGGGCCTGGATGACGAGGGCAGCAGCCAGGGCGAGCCACAGTCAAAGAGCCCCC 6111

Qy        480 LeuAsnMetLeu-AlaLysHisTyrAlaVal------AlaIleGlyCysGluIleValPh 497
Db       6112 AGGAGTCACGCCGGCTGAGCATCCAGCGCTGCATCCAGTCGCTGGTGCACGCGTGCCA-- 6169

Qy        497 ePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSe 517
Db       6170 ----GTGCCGCAACGC----------------------CAACTGCTCGCTGCCATCCT 6201

Qy        517 rSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValPr 537
Db       6202 GCCAGAAGATGAAGCGGGTGGTGCAGCACACCAAGGG---------------------- 6238

Qy        537 oProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyVa 557
Db       6239 ------------------------CTGCAAACGCAAGACCAACGGGGGCTGCCCGGTGT 6273

Qy        557 l--------------ValLeuLysLeuProProProAlaValProValThrAlaSerTy 572
Db       6274 GCAAGCAGCTCATCGCCCTCTGCTGCTACCACGCCAAGCACTGCCAAGAAAACAAATGCC 6333

Qy        572 rArg--------------GlnGluSerGlnSerAsnProLeuHisTyrMetAspGlnAl 587
Db       6334 CCGTGCCCTTCTGCCTCAACATCAAACACAAGCTCCGCCAGCAGCAGATCCAGCACCGCC 6393

Qy        587 aArgAspSerProAlaAsnAla---------SerHisSerLeuTyrProProArgGluAs 604
Db       6394 TGCAGCAGGCCCAGCTCATGCGCCGGCGGATGGCCACCATGAACACCCGCAACGTGCCTC 6453

Qy        604 nTyrIleArgGlyAlaProGluHisLeuThrAlaAlaSerLysProSerValSerGluPr 624
Db       6454 A---GCAGAGTCTGCCTTCTCCTACCTCAGCACCGCCCGGGACCCCCACACAGCAGCCCA 6510

Qy        624 oLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuLe 644
Db       6511 G---CACACCCCAGACGCCGCAGCCCCCTGCCCAGCCCCAACCCTCACCCGTGAG----- 6562

Qy        644 uAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro----------- 660
Db       6563 ----CATGTCACCAGCTGGCTTCCCCAGCGTGGCCCGGACTCAGCCCCCCACCACGGTGT 6618

Qy        661 -------GluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGl 678
Db       6619 CCACAGGGAAGCCTACCAGCCAGGTGCCGGCCCCCCCACC-------------------- 6658

Qy        678 nSerAsnGlyLeuTyrAsnGlyGluAlaProSer--GlnAlaTrpLysArgGlyProGln 697
Db       6659 ---------------CCCGGCCCAGCCCCCTCCTGCAGCGGTGGAAGCGGCT-CGGCAG 6701

Qy        698 ThrValHisAspAlaSerAsnGlnSer------PheGlnGlnTyrGlyAsnGlnTyrThr 715
Db       6702 ATCGAGCGTGAGGCCCAGCAGCAGCACCTGTACCGGGTGAACATCAACAACAGCATG 6761

Qy        716 ProAlaGlyGln------LeuProProProProSerArgTyrProProAlaSerAsnAsn 733
Db       6762 CCCCCAGGACGCACGGGCATGGGGACCCCGGGGAGCCAGATGGCCCCCGTGAGCCTGAAT 6821
```

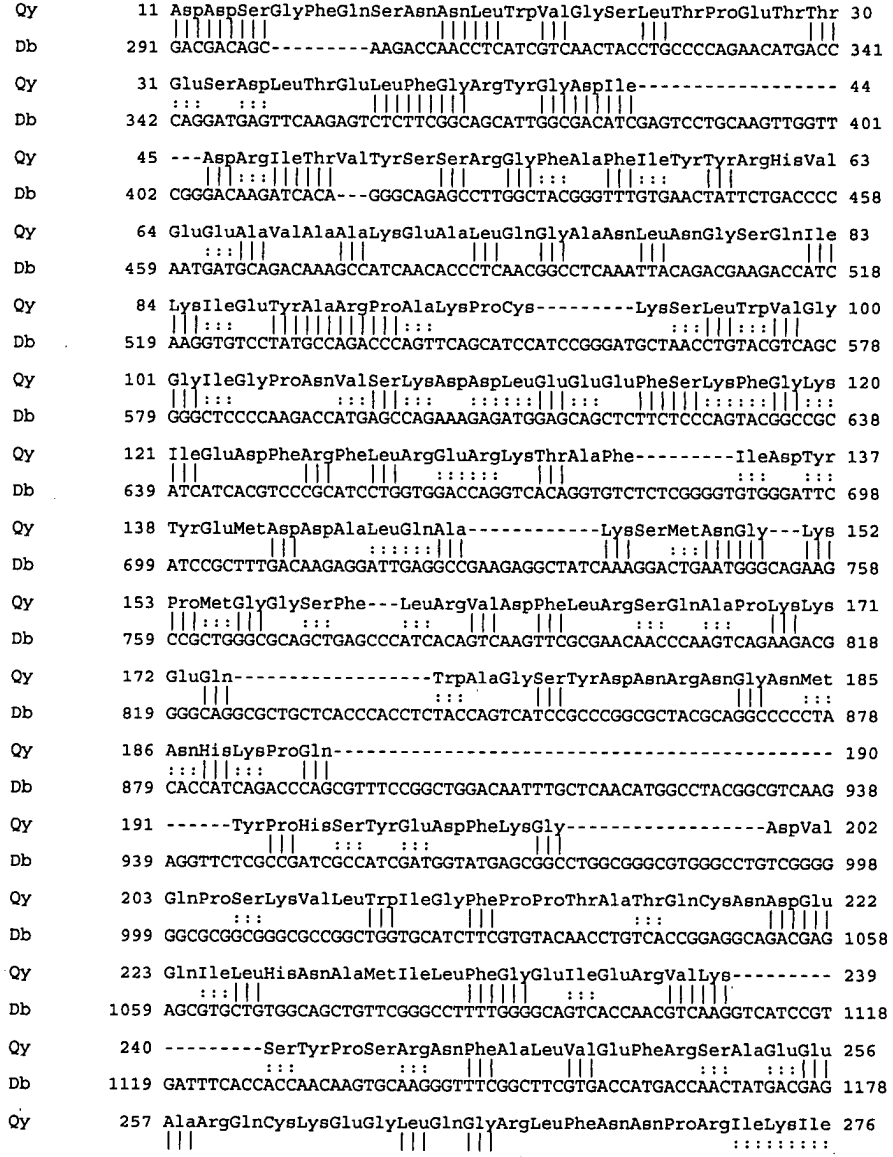


RESULT 13
US-09-949-016-3886
; Sequence 3886, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3886
; LENGTH: 1968
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-3886

Alignment Scores:

Pred. No.:	6.22e-08	Length:	1968
Score:	215.50	Matches:	101
Percent Similarity:	36.95%	Conservative:	59
Best Local Similarity:	23.33%	Mismatches:	178
Query Match:	4.49%	Indels:	96
DB:	4	Gaps:	17

US-09-920-705-3 (1-901) x US-09-949-016-3886 (1-1968)



```
Db       1179 GCGGCCATGGCCATCGCCAGCCTGAACGGCTATCGCCTGGGCGAGCGCGTGCTGCAGGTC 1238

Qy        277 MetTyrSerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMet 296
Db       1239 TCCTTC------------------------------------------------------ 1244

Qy        297 LysArgSerArgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSer 316
Db       1245 AAGACCAGCAAACAGCACAAGGCGTGAGCCCACCCCGCCTGCCC-TCCCACCCC------ 1297

Qy        317 ThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGly 336
Db       1298 ------CTCCCCGGGCAGCAGAGAGAGAGAGAGAGAAAGAGAGAGAGAGAGAGAAGGG 1351

Qy        337 AlaGluTyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGly 356
Db       1352 GCCCAAGAGAGACAG------CACAGGCAGCCCCACGGACGACGC-------------- 1390

Qy        357 ThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThr 376
Db       1391 --------------------GAGGGCCCCACGTCCCTGCGGAAGCCA---CAGGGTGAG 1426

Qy        377 ArgArgProMetArgSerAsnProAspSerTrpGluGly 389
Db       1427 CACTCTGGGGTGGGAGGGTCTGCAGGGAATTGGGGGGGT 1465

RESULT 14
US-09-410-551B-26
; Sequence 26, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
;  APPLICANT: KOSAN BIOSCIENCES, Inc.
;  APPLICANT:  REEVES, CHRISTOPHER
;  APPLICANT:  CHU, DANIEL
;  APPLICANT:  KHOSLA, CHAITAN
;  APPLICANT:  SANTI, DANIEL
;  APPLICANT:  WU, KAI
;  TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
;  TITLE OF INVENTION:  CONSTRUCTS THEREFOR
;  FILE REFERENCE: 30062-20026.00
;  CURRENT APPLICATION NUMBER: US/09/410,551B
;  CURRENT FILING DATE:  1999-10-01
;  PRIOR APPLICATION NUMBER: US 60/139,650
;  PRIOR FILING DATE: 1999-06-17
;  PRIOR APPLICATION NUMBER: US 60/123,810
;  PRIOR FILING DATE: 1999-03-11
;  PRIOR APPLICATION NUMBER: US 60/102,748
;  PRIOR FILING DATE: 1998-10-02
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
;   LENGTH: 4674
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
;   OTHER INFORMATION: PKS synthase fragment
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3)..(4673)
US-09-410-551B-26

Alignment Scores:
Pred. No.:              1.01e-06      Length:         4674
Score:                  208.00        Matches:        209
Percent Similarity:     31.27%        Conservative:   109
Best Local Similarity:  20.55%        Mismatches:     400
Query Match:            4.33%         Indels:         300
DB:                     4             Gaps:           44

US-09-920-705-3 (1-901) x US-09-410-551B-26 (1-4674)

Qy         10 AlaAspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThr 29

Db       1743 GCGGACAACGCGGTGATCGAGCGGGCACCGGAGTGGGTGCCGTTGGTGATTTCGGCCAGG 1802

Qy         30 ThrGluSerAspLeuThrGluLeuPheGlyArgTyr--------------------Gly 42
Db       1803 ACCCAGTCGGCTTTGACTGAGCACGAGGGCCGGTTGCGTGCGTATCTGGCGGCGTCGCCC 1862

Qy         43 AspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyrArgHis 62
Db       1863 GGGGTGGATATGCGGGCTGTGGCATCGACGCTGGCGATGACACGGTCGGTGTTCGAGCAC 1922

Qy         63 ------------------------------ValGluGluAlaValAlaAlaLysGluAla 72
Db       1923 CGTGCCGTGCTGCTGGGAGATGACACCGTCACCGGCACCGCTGTGTCTGACCCTCGGGCG 1982

Qy         73 Leu---------------GlnGlyAlaAsnLeuAsnGlySerGlnIleLysIleGluTyr 87
Db       1983 GTGTTCGTCTTCCCGGGACAGGGGTCGCAGCGTGCTGGCATGGGTGAGGAACTGGCCGCC 2042

Qy         88 AlaArgPro------AlaLysProCysLysSerLeuTrpValGlyGlyIleGlyProAsn 105
Db       2043 GCGTTCCCCGTCTTCGCGCGGATCCATCAGCAGGTGTGG--------------GACCTG 2087

Qy        106 ValSerLysAspAspLeuGlu------------------------------------- 112
Db       2088 CTCGATGTGCCCGATCTGGAGGTGAACGAGACCGGTTACGCCCAGCCGGCCCTGTTCGCA 2147

Qy        113 GluGluPheSerLysPheGlyLysIleGluAspPhe---------------------- 124
Db       2148 ATGCAGGTGGCTCTGTTCGGGCTGCTGGAATCGTGGGGTGTACGACCGGACGCGGTGATC 2207

Qy        125 ---ArgPheLeuArgGluArgLysThrAlaPheIleAsp---TyrTyrGluMetAspAsp 142
Db       2208 GGCCATTCGGTGGGTGAGCTTGCGGCTGCGTATGTGTCCGGGGTGTGGTCGTTGGAGGAT 2267

Qy        143 Ala---------------LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySer 157
Db       2268 GCCTGCACTTTGGTGTCGGCGCGGGCTCGTCTGATGCAGGCTCTGCCCGCGGGTGGGGTG 2327

Qy        158 PheLeuArgValAspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySer 177
Db       2328 ATGGTCGCTGTC------------------------------------------------ 2339

Qy        178 TyrAspAsnArgAsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAsp 197
Db       2340 ----------------------------------------CCGGTCTCGGAGGATGAG 2357

Qy        198 PheLysGlyAspVal---GlnProSerLysValLeuTrpIleGlyPheProProThrAla 216
Db       2358 GCCCGGGCCGTGCTGGGTGAGGGTGTGGAGATCGCCGCGGTCAACGGCCCGTCGTCGGTG 2417

Qy        217 ThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGlu 236
Db       2418 GTTCTCTCCGGTGATGAGGCCGCCGTGCTGCAGGCCGCGGAGGGGCTGGGGAAGTGGACG 2477

Qy        237 ArgValLysSerTyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGlu 256
Db       2478 CGGCTGGCGACCAGCCACGCGTTCCATTCCGCCCGTATGGAA---CCCATGCTGGAGGAG 2534

Qy        257 AlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIle 276
Db       2535 TTCCGGGCGGTCGCCGAAGGCCTG---------ACCTACCGGACGCCGCAGGTCTCCATG 2585

Qy        277 MetTyrSerAsnAspGluLeuProProGlu------------------GlnAspAspThr 290
Db       2586 GCCGTTGGTGATCAGGTGACCACCGCTGAGTACTGGGTGCGGCAGGTCCGGGACACGGTC 2645

Qy        291 SerPhe-TyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsn----------- 306
Db       2646 CGGTTCGGCGAGCAGGTGGCCTCGTACGAGGACGCCGTGTTCGTCGAGCTGGGTGCCGAC 2705

Qy        306 ------------------------------------------------------------ 306
```

```
Db       2706 CGGTCACTGGCCCGCCTGGTCGACGGTGTCGCGATGCTGCACGGCGACCACGAAATCCAG 2765
Qy        307 ----------------AspProSerCysValSer------------------------- 312
Db       2766 GCCGCGATCGGCGCCCTGGCCCACCTGTATGTCAACGGCGTCACGGTCGACTGGCCCGCG 2825
Qy        313 ----------------------------------------------SerProHisSerTh 317
Db       2826 CTCCTGGGCGATGCTCCGGCAACACGGGTGCTGGACCTTCCGACATACGCCTTCCAGCAC 2885
Qy        317 rGlyIleProGlySerMet---------------ArgProLeuArgGlyThrAsnGluAr 332
Db       2886 CAGCGCTACTGGCTCGAGTCGGCTCCCCCGGCCACGGCCGACTCGGGCCACCCCGTCCTC 2945
Qy        332 gSerTyrAsnGlyAlaGluTyrAsnAspValValGly----------------------- 344
Db       2946 GGCACCGGAGTCGCCGTCGCCGGGTCGCCGGGCCGGGTGTTCACGGGTCCCGTGCCCGCC 3005
Qy        345 -------------------LysGluProAsnTrpArg----------------------- 350
Db       3006 GGTGCGGACCGCGCGGTGTTCATCGCCGAACTGGCGCTCGCCGCCGCCGACGCCACCGAC 3065
Qy        351 -ArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuPr 370
Db       3066 TGCGCCACGGTCGAACAGCTCGACGTCACCTCCGTGCCCGGCGGATCCG------------ 3114
Qy        370 oSerProAlaGlnGly---ThrArgArgProMetArgSerAsnProAspSerTrpGluGl 389
Db       3115 ----CCCGCGGCAGGGCCACCGCGCAGACCTGGGTCGATGAACCCGCCGCCGACGGGCGG 3170
Qy        389 y------------TyrAspProAlaGlnLeuValArgGluSerLysArgThrArgArgAs 405
Db       3171 CGCCGCTTCACCGTCCACACCCGCGTCGGCGACGCCCCGTGGACGCTGCACGCCGAGGGG 3230
Qy        405 pGlySerValAspGly--------------------PheThrProMetGlyValAspGl 418
Db       3231 GTTCTCCGCCCCGGCCGCGTGCCCCAGCCCGAAGCCGTCGACACCGCCTGGCCCCCGCCG 3290
Qy        418 uArgSerPhe------------------GlyArgGlySerValAlaAlaArgProIleAr 432
Db       3291 GGCGCGGTGCCCGCGGACGGGCTGCCCGGGGCGTGGCGACGCGCGGACCAGGTCTTCGTC 3350
Qy        432 gGlyProProAspSerAspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrProVa 452
Db       3351 GAAGCCGAAGTCGACA--------------GCCCTGACGGCTTCGTGGCACACCCCGAC 3395
Qy        452 lCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluValVa 472
Db       3396 CTGCTCGACGCGGTCTTCTCCGCGG-------------------------TCGGCGAC 3428
Qy        472 lAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGl 492
Db       3429 GGGAGCCGCCAGCCGACCGGATGG-CGCGACCTCGCGGTGCAC---------------- 3470
Qy        492 yCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPh 512
Db       3471 ----------------------------------------GCGTCGGACGCCACCGT 3487
Qy        512 eLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAsp------AspGl 530
Db       3488 GCTGCGCGCCTGCCTCACCCGCCGCGACAGTGGTGTCGTGGAGCTCGCCGCCTTCGACGG 3547
Qy        530 yThrThrLeu---PheLeuValProProSerAspPheLeuThrAspValLeuGlnValTh 549
Db       3548 TGCCGGAATGCCGGTGCTCACCGCGGAGTCGGTGACGCTGGGCGAGGTCGCGTCGGCAGG 3607
Qy        549 rArgGlnGluArgLeuTyrGlyValVal------------LeuLysLeuProPro-ProA 565
Db       3608 CGGATCCGACGAGTCGGACGGTCTGCTTCGGCTTGAGTGGTTGCCGGTGGCGGAGGCCCA 3667
Qy        565 laValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnProLeuHisTyrMetA 585
Db       3668 CTACGACGGTGCCGACGAGCTGCCCGAGGGCTACACCCTCATCACCGCCACACACCCCGA 3727

Qy        585 spGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGluAsnT 605
Db       3728 CGACCCCGACGACCCCACCAACCCCCACAACACACCCACACGCACCCACACACAAACCAC 3787
Qy        605 yrIleArgGlyAlaProGluHisLeuThrAlaAla------SerLysProSerValSerG 623
Db       3788 ACGCGTCCTCACCGCCCTCCAACACCACCTCATCACCACCAACCACACCCTCATCGTCCA 3847
Qy        623 luProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluL 643
Db       3848 CACCA------CCACCGACCCCCCAGGCGCCGCCGTCACCGGCCTCACCCGCACCGCACA 3901
Qy        643 euLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerH 663
Db       3902 AAACGAACACCCCGGCCGCA------TCCACCTCATCGAAACCCACCACCCCCACACCCC 3955
Qy        663 isGlnProMet------SerGlyProSerThr-ValValSerThrAlaHisGlnSerAsn 680
Db       3956 ACTCCCCCTCACCCAACTCACCACCCTCCACCAACCCCACCTACGCCTCACCAACAA--- 4012
Qy        681 GlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHis 700
Db       4013 ------CACCCTCCACACCCCCCACCTCACCCCCATCACCACCCACCACAACACCACCAC 4066
Qy        701 AspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeu 720
Db       4067 AAC-------------------------------------CACCCCCAACACCCCACC 4087
Qy        721 ProProProProSerArgTyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMet 740
Db       4088 CCTCAACCCCAACCACGCCATCCTCAT------------------------------- 4114
Qy        741 ValHisGlyAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerPro--- 759
Db       4115 ---CACCGGCGGCTCCGGCACCCTCGCCGGCATCCTCGCCCGCCACCTCAACCACCCCCA 4171
Qy        760 ---LeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHis 778
Db       4172 CACCTACCTCCTCTCCCGCACACCACCACCCCCCACCACACCCGG----------CAC 4219
Qy        779 ProValSerGlnProMetValGlnGlnTyrGlnProGluAlaSerMetProAsnGlnAsn 798
Db       4220 CCA------CATCCCCTG---CGACCTCACCGACCCCACCCAAATCACCCAAGCCCTCAC 4270
Qy        799 TyrGlyPro-IleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThr-------- 815
Db       4271 ------CCACATACCACAACCCCTCACCGGCATCTTCCACACCGCCGCCACCCTCGACGA 4324
Qy        816 -AsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAs 835
Db       4325 CGCCACCCTCACCAACCTCACCCCCCAACACCTCACCACCACCCTCCAACCCAAAGCCGA 4384
Qy        835 pLysAla---------AsnLeuGluProGlnAsnGlnAlaLeu 846
Db       4385 CGCCGCCTGGCACCTCCACCACCACACCCAAAACCAACCCCTC 4427
```

RESULT 15
US-09-940-316B-26
; Sequence 26, Application US/09940316B
; Patent No. 6759536
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTH
; TITLE OF INVENTION: GENE CLUSTER
; FILE REFERENCE: 30062-20026.11
; CURRENT APPLICATION NUMBER: US/09/940,316B
; CURRENT FILING DATE: 2001-08-27

```
;  PRIOR APPLICATION NUMBER: 09/410,551
;  PRIOR FILING DATE: 1999-10-01
;  PRIOR APPLICATION NUMBER: US 60/139,650
;  PRIOR FILING DATE: 1999-06-17
;  PRIOR APPLICATION NUMBER: US 60/123,810
;  PRIOR FILING DATE: 1999-03-11
;  PRIOR APPLICATION NUMBER: US 60/102,748
;  PRIOR FILING DATE: 1998-10-02
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
;   LENGTH: 4674
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
;   OTHER INFORMATION: PKS synthase fragment
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3)..(4673)
US-09-940-316B-26

Alignment Scores:
Pred. No.:                1.01e-06        Length:          4674
Score:                    208.00          Matches:         209
Percent Similarity:       31.27%          Conservative:    109
Best Local Similarity:    20.55%          Mismatches:      400
Query Match:              4.33%           Indels:          300
DB:                       4               Gaps:            44

US-09-920-705-3 (1-901) x US-09-940-316B-26 (1-4674)

Qy         10 AlaAspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThr 29
Db       1743 GCGGACAACGCGGTGATCGAGCGGGCACCGGAGTGGGTGCCGTTGGTGATTTCGGCCAGG 1802

Qy         30 ThrGluSerAspLeuThrGluLeuPheGlyArgTyr----------------------Gly 42
Db       1803 ACCCAGTCGGCTTTGACTGAGCACGAGGGCCGGTTGCGTGCGTATCTGGCGGCGTCGCCC 1862

Qy         43 AspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyrArgHis 62
Db       1863 GGGGTGGATATGCGGGCTGTGGCATCGACGCTGGCGATGACACGGTCGGTGTTCGAGCAC 1922

Qy         63 ----------------------------ValGluGluAlaValAlaAlaLysGluAla 72
Db       1923 CGTGCCGTGCTGCTGGGAGATGACACCGTCACCGGCACCGCTGTGTCTGACCCTCGGGCG 1982

Qy         73 Leu---------------GlnGlyAlaAsnLeuAsnGlySerGlnIleLysIleGluTyr 87
Db       1983 GTGTTCGTCTTCCCGGGACAGGGGTCGCAGCGTGCTGGCATGGGTGAGGAACTGGCCGCC 2042

Qy         88 AlaArgPro------AlaLysProCysLysSerLeuTrpValGlyGlyIleGlyProAsn 105
Db       2043 GCGTTCCCCGTCTTCGCGCGGATCCATCAGCAGGTGTGG---------------GACCTG 2087

Qy        106 ValSerLysAspAspLeuGlu------------------------------------- 112
Db       2088 CTCGATGTGCCCGATCTGGAGGTGAACGAGACCGGTTACGCCCAGCCGGCCCTGTTCGCA 2147

Qy        113 GluGluPheSerLysPheGlyLysIleGluAspPhe---------------------- 124
Db       2148 ATGCAGGTGGCTCTGTTCGGGCTGCTGGAATCGTGGGGTGTACGACCGGACGCGGTGATC 2207

Qy        125 ---ArgPheLeuArgGluArgLysThrAlaPheIleAsp---TyrTyrGluMetAspAsp 142
Db       2208 GGCCATTCGGTGGGTGAGCTTGCGGCTGCGTATGTGTCCGGGGTGTGGTCGTTGGAGGAT 2267

Qy        143 Ala---------------LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySer 157
Db       2268 GCCTGCACTTTGGTGTCGGCGCGGGCTCGTCTGATGCAGGCTCTGCCCGCGGGTGGGGTG 2327

Qy        158 PheLeuArgValAspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySer 177

Db       2328 ATGGTCGCTGTC---------------------------------------------- 2339

Qy        178 TyrAspAsnArgAsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAsp 197
Db       2340 ----------------------------------------CCGGTCTCGGAGGATGAG 2357

Qy        198 PheLysGlyAspVal---GlnProSerLysValLeuTrpIleGlyPheProProThrAla 216
Db       2358 GCCCGGGCCGTGCTGGGTGAGGGTGTGGAGATCGCCGCGGTCAACGGCCCGTCGTCGGTG 2417

Qy        217 ThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGlu 236
Db       2418 GTTCTCTCCGGTGATGAGGCCGCCGTGCTGCAGGCCGCGGAGGGGCTGGGGAAGTGGACG 2477

Qy        237 ArgValLysSerTyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGlu 256
Db       2478 CGGCTGGCGACCAGCCACGCGTTCCATTCCGCCCGTATGGAA---CCCATGCTGGAGGAG 2534

Qy        257 AlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIle 276
Db       2535 TTCCGGGCGGTCGCCGAAGGCCTG---------ACCTACCGGACGCCGCAGGTCTCCATG 2585

Qy        277 MetTyrSerAsnAspGluLeuProProGlu------------------GlnAspAspThr 290
Db       2586 GCCGTTGGTGATCAGGTGACCACCGCTGAGTACTGGGTGCGGCAGGTCCGGGACACGGTC 2645

Qy        291 SerPhe-TyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsn----------- 306
Db       2646 CGGTTCGGCGAGCAGGTGGCCTCGTACGAGGACGCCGTGTTCGTCGAGCTGGGTGCCGAC 2705

Qy        306 ------------------------------------------------------------ 306
Db       2706 CGGTCACTGGCCCGCCTGGTCGACGGTGTCGCGATGCTGCACGGCGACCACGAAATCCAG 2765

Qy        307 ----------------AspProSerCysValSer-------------------------- 312
Db       2766 GCCGCGATCGGCGCCCTGGCCCACCTGTATGTCAACGGCGTCACGGTCGACTGGCCCGCG 2825

Qy        313 ------------------------------------------SerProHisSerTh 317
Db       2826 CTCCTGGGCGATGCTCCGGCAACACGGGTGCTGGACCTTCCGACATACGCCTTCCAGCAC 2885

Qy        317 rGlyIleProGlySerMet--------------ArgProLeuArgGlyThrAsnGluAr 332
Db       2886 CAGCGCTACTGGCTCGAGTCGGCTCCCCCGGCCACGGCCGACTCGGGCCACCCCGTCCTC 2945

Qy        332 gSerTyrAsnGlyAlaGluTyrAsnAspValValGly---------------------- 344
Db       2946 GGCACCGGAGTCGCCGTCGCCGGGTCGCCGGGCCGGGTGTTCACGGGTCCCGTGCCCGCC 3005

Qy        345 -------------------LysGluProAsnTrpArg---------------------- 350
Db       3006 GGTGCGGACCGCGCGGTGTTCATCGCCGAACTGGCGCTCGCCGCCGCCGACGCCACCGAC 3065

Qy        351 -ArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuPr 370
Db       3066 TGCGCCACGGTCGAACAGCTCGACGTCACCTCCGTGCCCGGCGGATCCG----------- 3114

Qy        370 oSerProAlaGlnGly---ThrArgArgProMetArgSerAsnProAspSerTrpGluGl 389
Db       3115 ----CCCGCGGCAGGGCCACCGCGCAGACCTGGGTCGATGAACCCGCCGCCGACGGGCGG 3170

Qy        389 y------------TyrAspProAlaGlnLeuValArgGluSerLysArgThrArgArgAs 405
Db       3171 CGCCGCTTCACCGTCCACACCCGCGTCGGCGACGCCCCGTGGACGCTGCACGCCGAGGGG 3230

Qy        405 pGlySerValAspGly--------------------PheThrProMetGlyValAspGl 418
Db       3231 GTTCTCCGCCCCGGCCGCGTGCCCCAGCCCGAAGCCGTCGACACCGCCTGGCCCCCGCCG 3290

Qy        418 uArgSerPhe-----------------GlyArgGlySerValAlaAlaArgProIleAr 432
```

```
Db       3291 GGCGCGGTGCCCGCGGACGGGCTGCCCGGGGCGTGGCGACGCGCGGACCAGGTCTTCGTC 3350

Qy        432 gGlyProProAspSerAspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrProVa 452

Db       3351 GAAGCCGAAGTCGACA--------------GCCCTGACGGCTTCGTGGCACACCCCGAC 3395

Qy        452 lCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluValVa 472

Db       3396 CTGCTCGACGCGGTCTTCTCCGCGG--------------------------TCGGCGAC 3428

Qy        472 lAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGl 492

Db       3429 GGGAGCCGCCAGCCGACCGGATGG-CGCGACCTCGCGGTGCAC---------------- 3470

Qy        492 yCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPh 512

Db       3471 ----------------------------------------GCGTCGGACGCCACCGT 3487

Qy        512 eLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAsp------AspGl 530

Db       3488 GCTGCGCGCCTGCCTCACCCGCCGCGACAGTGGTGTCGTGGAGCTCGCCGCCTTCGACGG 3547

Qy        530 yThrThrLeu---PheLeuValProProSerAspPheLeuThrAspValLeuGlnValTh 549

Db       3548 TGCCGGAATGCCGGTGCTCACCGCGGAGTCGGTGACGCTGGGCGAGGTCGCGTCGGCAGG 3607

Qy        549 rArgGlnGluArgLeuTyrGlyValVal----------LeuLysLeuProPro-ProA 565

Db       3608 CGGATCCGACGAGTCGGACGGTCTGCTTCGGCTTGAGTGGTTGCCGGTGGCGGAGGCCCA 3667

Qy        565 laValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnProLeuHisTyrMetA 585

Db       3668 CTACGACGGTGCCGACGAGCTGCCCGAGGGCTACACCCTCATCACCGCCACACACCCCGA 3727

Qy        585 spGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGluAsnT 605

Db       3728 CGACCCCGACGACCCCACCAACCCCCACAACACACCCACACGCACCCACACACAAACCAC 3787

Qy        605 yrIleArgGlyAlaProGluHisLeuThrAlaAla------SerLysProSerValSerG 623

Db       3788 ACGCGTCCTCACCGCCCTCCAACACCACCTCATCACCACCAACCACACCCTCATCGTCCA 3847

Qy        623 luProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluL 643

Db       3848 CACCA------CCACCGACCCCCCAGGCGCCGCCGTCACCGGCCTCACCCGCACCGCACA 3901

Qy        643 euLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerH 663

Db       3902 AAACGAACACCCCGGCCGCA------TCCACCTCATCGAAACCCACCACCCCCACACCCC 3955

Qy        663 isGlnProMet------SerGlyProSerThr-ValValSerThrAlaHisGlnSerAsn 680

Db       3956 ACTCCCCCTCACCCAACTCACCACCCTCCACCAACCCCACCTACGCCTCACCAACAA--- 4012

Qy        681 GlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHis 700

Db       4013 ------CACCCTCCACACCCCCCACCTCACCCCCATCACCACCCACCACAACACCACCAC 4066

Qy        701 AspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeu 720

Db       4067 AAC---------------------------------------CACCCCCAACACCCCACC 4087

Qy        721 ProProProProSerArgTyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMet 740

Db       4088 CCTCAACCCCAACCACGCCATCCTCAT------------------------------- 4114

Qy        741 ValHisGlyAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerPro--- 759

Db       4115 ---CACCGGCGGCTCCGGCACCCTCGCCGGCATCCTCGCCCGCCACCTCAACCACCCCCA 4171

Qy        760 ---LeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHis 778

Db       4172 CACCTACCTCCTCTCCCGCACACCACCACCCCCCACCACACCCGG-----------CAC 4219

Qy        779 ProValSerGlnProMetValGlnGlnTyrGlnProGluAlaSerMetProAsnGlnAsn 798

Db       4220 CCA------CATCCCCTG---CGACCTCACCGACCCCACCCAAATCACCCAAGCCCTCAC 4270

Qy        799 TyrGlyPro-IleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThr-------- 815

Db       4271 ------CCACATACCACAACCCCTCACCGGCATCTTCCACACCGCCGCCACCCTCGACGA 4324

Qy        816 -AsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAs 835

Db       4325 CGCCACCCTCACCAACCTCACCCCCCAACACCTCACCACCACCCTCCAACCCAAAGCCGA 4384

Qy        835 pLysAla---------AsnLeuGluProGlnAsnGlnAlaLeu 846

Db       4385 CGCCGCCTGGCACCTCCACCACCACACCCAAAACCAACCCCTC 4427
```

Search completed: April 8, 2005, 13:52:02
Job time : 453 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2005, 12:01:35 ; Search time 1224 Seconds
(without alignments)
4460.836 Million cell updates/sec

Title: US-09-920-705-3
Perfect score: 4801
Sequence: 1 MALSMKPFRADDSGFQSNNL.....LQIQQKQQQQSSGTPAGQGP 901

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5615251 seqs, 3030001701 residues

Total number of hits satisfying chosen parameters: 11230502

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09920705/runat_07042005_101816_20810/app_query.fasta_1.1095
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09920705_@CGN_1_1_622_@runat_07042005_101816_20810
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4801	100.0	2706	10	US-09-920-705-2	Sequence 2, Appli
2	4392	91.5	4593	10	US-09-920-705-1	Sequence 1, Appli
c 3	1436	29.9	801	10	US-09-920-705-4	Sequence 4, Appli
4	1286.5	26.8	3463	18	US-10-425-115-123950	Sequence 123950,
5	1281.5	26.7	3084	18	US-10-437-963-36321	Sequence 36321, A
6	802	16.7	3315	18	US-10-437-963-79215	Sequence 79215, A
7	669	13.9	1101	17	US-10-424-599-59704	Sequence 59704, A
8	651.5	13.6	3715	10	US-09-920-705-6	Sequence 6, Appli
9	499	10.4	801	17	US-10-424-599-65572	Sequence 65572, A
10	466	9.7	349	10	US-09-920-705-5	Sequence 5, Appli
11	463	9.6	1630	17	US-10-425-114-29888	Sequence 29888, A
12	447	9.3	1631	17	US-10-424-599-40095	Sequence 40095, A
13	406	8.5	2159	17	US-10-425-114-31541	Sequence 31541, A
14	382	8.0	815	18	US-10-767-701-2544	Sequence 2544, Ap
15	359	7.5	606	18	US-10-767-701-30057	Sequence 30057, A
16	336.5	7.0	6836	18	US-10-475-917-1	Sequence 1, Appli
17	333.5	6.9	1146	17	US-10-424-599-18149	Sequence 18149, A
18	309	6.4	729	17	US-10-424-599-47401	Sequence 47401, A
19	302	6.3	483728	18	US-10-699-156-2	Sequence 2, Appli
20	297	6.2	1621	18	US-10-425-115-118399	Sequence 118399,
21	276.5	5.8	3383	18	US-10-475-917-11	Sequence 11, Appl
22	273	5.7	3312	18	US-10-475-917-7	Sequence 7, Appli
23	273	5.7	3423	18	US-10-475-917-9	Sequence 9, Appli
24	271.5	5.7	10656	17	US-10-263-929-42	Sequence 42, Appl
25	270	5.6	1241	17	US-10-425-114-35702	Sequence 35702, A
26	268	5.6	3228	17	US-10-302-172-752	Sequence 752, App
27	265	5.5	10995	17	US-10-263-929-41	Sequence 41, Appl
28	265	5.5	12227	15	US-10-177-293-422	Sequence 422, App
29	265	5.5	12227	18	US-10-684-422-243	Sequence 243, App
30	261	5.4	11295	10	US-09-814-353-19153	Sequence 19153, A
31	258.5	5.4	683	17	US-10-424-599-2550	Sequence 2550, Ap
32	254	5.3	438	9	US-09-770-444-972	Sequence 972, App
33	246.5	5.1	1871	9	US-09-834-975-1036	Sequence 1036, Ap
34	246	5.1	2228	17	US-10-172-118-410	Sequence 410, App
35	246	5.1	2228	17	US-10-342-887-410	Sequence 410, App
36	246	5.1	2482	18	US-10-723-860-6794	Sequence 6794, Ap
37	241.5	5.0	443	18	US-10-425-115-79600	Sequence 79600, A
38	236.5	4.9	8147	13	US-10-109-886-9	Sequence 9, Appli
39	229.5	4.8	1803	9	US-09-938-842A-1801	Sequence 1801, Ap
40	229.5	4.8	1803	11	US-09-938-842A-1801	Sequence 1801, Ap
41	225	4.7	16683	17	US-10-263-929-40	Sequence 40, Appl
42	223.5	4.7	2201	18	US-10-425-115-118394	Sequence 118394,
43	218.5	4.6	2340	18	US-10-425-115-31949	Sequence 31949, A
44	216	4.5	2774	18	US-10-723-860-8366	Sequence 8366, Ap
c 45	212	4.4	3433	18	US-10-425-115-25600	Sequence 25600, A

ALIGNMENTS

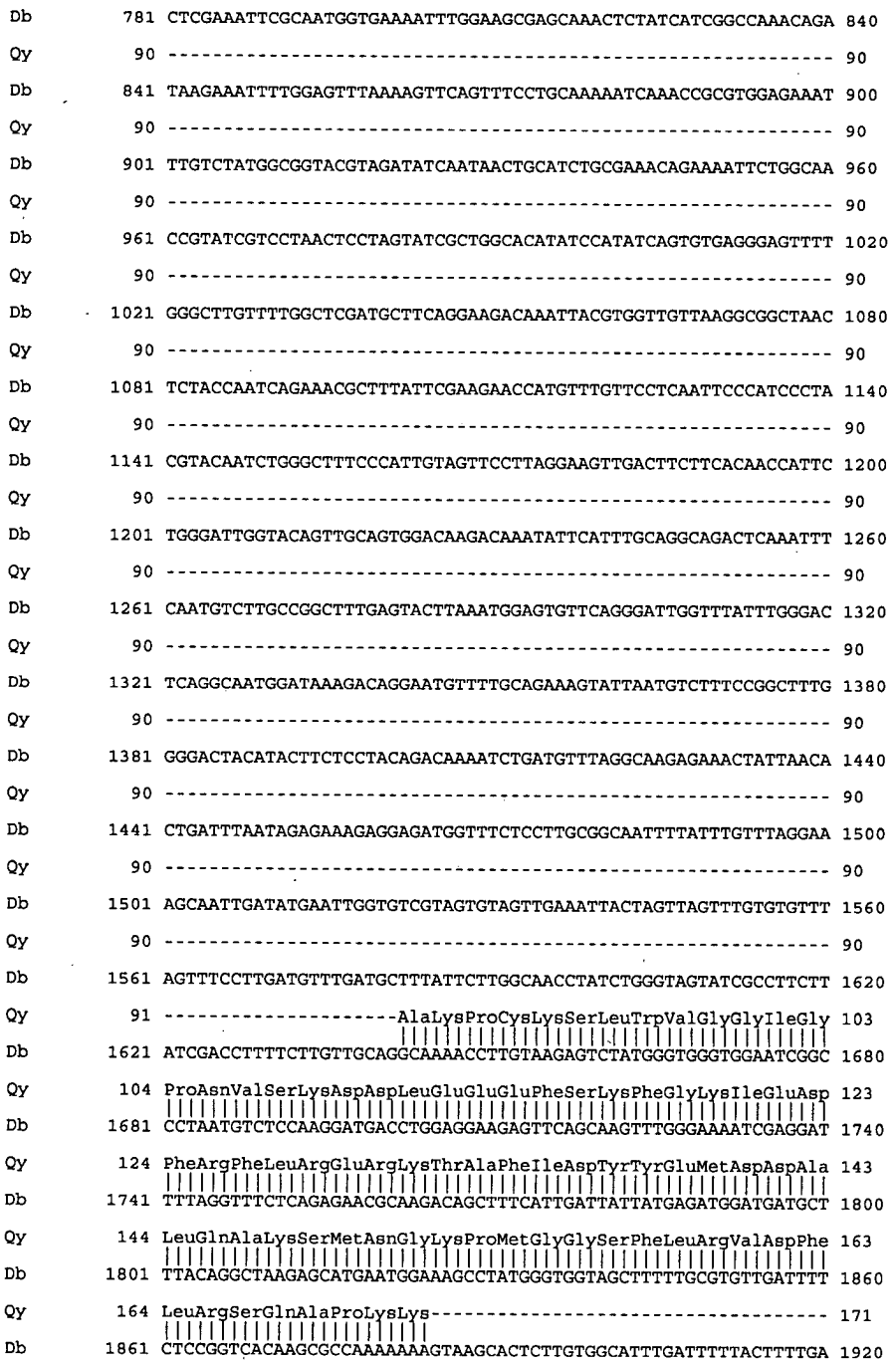
RESULT 1
US-09-920-705-2
; Sequence 2, Application US/09920705
; Publication No. US20030079252A1
; GENERAL INFORMATION:
; APPLICANT: Amasino, Richard M.
; APPLICANT: Schomburg, Fritz M.
; APPLICANT: Michaels, Scott D.
; APPLICANT: Patton, David
; TITLE OF INVENTION: Floral Induction Gene
; FILE REFERENCE: 960296.97214
; CURRENT APPLICATION NUMBER: US/09/920,705
; CURRENT FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2706
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:

```
;   NAME/KEY: CDS
;   LOCATION: (1)..(2706)
;   NAME/KEY: misc_binding
;   LOCATION: (46)..(279)
;   OTHER INFORMATION: RNA Binding Region
;   NAME/KEY: misc_binding
;   LOCATION: (283)..(522)
;   OTHER INFORMATION: RNA Binding Region
;   NAME/KEY: misc_binding
;   LOCATION: (610)..(852)
;   OTHER INFORMATION: RNA Binding Region
US-09-920-705-2

Alignment Scores:
Pred. No.:                0              Length:          2706
Score:                    4801.00        Matches:         901
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       10             Gaps:            0

US-09-920-705-3 (1-901) x US-09-920-705-2 (1-2706)

Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 60

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 120

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 180

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 240

Qy         81 SerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTCAAATTAAGATCGAATACGCACGACCGGCAAAACCTTGTAAGAGTCTATGGGTGGGT 300

Qy        101 GlyIleGlyProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAATCGGCCCTAATGTCTCCAAGGATGACCTGGAGGAAGAGTTCAGCAAGTTTGGGAAA 360

Qy        121 IleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMet 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATCGAGGATTTTAGGTTTCTCAGAGAACGCAAGACAGCTTTCATTGATTATTATGAGATG 420

Qy        141 AspAspAlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArg 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATGATGCTTTACAGGCTAAGAGCATGAATGGAAAGCCTATGGGTGGTAGCTTTTTGCGT 480

Qy        161 ValAspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsn 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTTGATTTTCTCCGGTCACAAGCGCCAAAAAAAGAACAATGGGCTGGCTCTTACGATAAC 540

Qy        181 ArgAsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGly 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGAAATGGCAATATGAATCATAAACCGCAGTATCCTCACTCATATGAAGACTTTAAAGGA 600

Qy        201 AspValGlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GATGTCCAGCCAAGTAAGGTTCTGTGGATTGGGTTCCCTCCTACTGCTACACAATGCAAT 660

Qy        221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATGAGCAAATTCTGCACAATGCGATGATACTCTTTGGTGAGATCGAGAGGGTAAAAAGT 720

Qy        241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        721 TACCCATCAAGGAATTTTGCACTTGTGGAGTTTAGGAGCGCGGAGGAAGCTCGCCAATGC 780

Qy        261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGGAAGGCCTACAGGGGAGGTTATTCAATAATCCTAGAATCAAAATTATGTACTCAAAC 840

Qy        281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GATGAGTTGCCTCCTGAGCAAGACGATACTAGTTTTTACTCTGGTATGAAACGGTCAAGG 900

Qy        301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACAGATATGTTCAATAATGATCCTTCATGTGTATCTTCTCCTCATTCTACTGGAATTCCT 960

Qy        321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsn 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGGTCTATGAGGCCCCTCAGAGGTACGAATGAGCGTTCATATAATGGTGCAGAATACAAT 1020

Qy        341 AspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GACGTTGTTGGTAAGGAGCCAAACTGGAGGAGGCCATCTGCAAATGGAACTGGAATACTC 1080

Qy        361 ProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMet 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCATCTCCAACAGGACCTGGAATCCTCCCATCTCCTGCACAAGGTACGAGGCGCCCTATG 1140

Qy        381 ArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuValArgGluSerLys 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AGGTCAAACCCCGATTCTTGGGAAGGATATGATCCTGCTCAGTTGGTCAGAGAAAGTAAA 1200

Qy        401 ArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGluArgSer 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGAACCAGAAGAGATGGATCAGTGGACGGTTTTACTCCAATGGGTGTCGATGAGAGGTCA 1260

Qy        421 PheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAspSerAspHisIle 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTTGGTCGAGGTTCAGTTGCTGCTAGACCTATCCGTGGCCCCCCTGATTCTGATCACATA 1320

Qy        441 TrpArgGlyMetIleAlaLysGlyGlyThrProValCysCysAlaArgCysValProMet 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGAGAGGAATGATTGCCAAGGGTGGAACTCCCGTCTGTTGTGCTCGTTGTGTACCTATG 1380

Qy        461 GlyLysGlyIleGluThrLysLeuProGluValValAsnCysSerAlaArgThrAspLeu 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GGAAAGGGGATTGAAACTAAACTGCCTGAGGTCGTCAATTGTTCAGCAAGAACTGATTTG 1440

Qy        481 AsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGluIleValPhePheValPro 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AATATGCTCGCTAAACATTACGCCGTTGCCATTGGATGTGAGATCGTTTTTTTCGTACCA 1500

Qy        501 AspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAsp 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GACAGGGAAGAAGATTTTGCGTCTTACACTGAATTTCTCCGGTACCTTAGCTCAAAAGAT 1560

Qy        521 ArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProProSerAsp 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CGGGCGGGTGTTGCCAAATTAGATGATGGTACAACTTTATTCTTGGTGCCTCCATCAGAT 1620

Qy        541 PheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLys 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTTAACTGATGTACTCCAAGTGACCCGTCAAGAACGGCTATATGGTGTTGTTCTCAAG 1680

Qy        561 LeuProProProAlaValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnPro 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTACCCCCGCCAGCCGTTCCTGTTACAGCATCATACAGACAAGAATCTCAGTCCAATCCT 1740

Qy        581 LeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrPro 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTGCATTATATGGATCAAGCCCGGGATTCACCTGCCAATGCTAGTCACAGTTTATATCCT 1800

Qy        601 ProArgGluAsnTyrIleArgGlyAlaProGluHisLeuThrAlaAlaSerLysProSer 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCTAGGGAAAATTACATTAGGGGTGCACCAGAACATTTGACAGCTGCTTCAAAACCATCT 1860
```

```
Qy        621 ValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThr 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GTTAGCGAGCCTCTCAGAATACCTAATAATGCAGCGCCTCAAGCTGGGGTTAGTTTAACT 1920

Qy        641 ProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCGGAGCTTTTAGCCACTCTGGCATCTATTCTCCCTGCAACTTCTCAACCTGCTGCCCCT 1980

Qy        661 GluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsn 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GAGAGTCACCAACCTATGTCAGGACCTTCAACAGTTGTTTCCACAGCACATCAGTCCAAT 2040

Qy        681 GlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHis 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GGACTGTACAATGGAGAAGCACCGTCTCAAGCTTGGAAAAGAGGTCCACAAACAGTTCAT 2100

Qy        701 AspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeu 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GATGCGTCAAATCAGTCATTCCAACAATACGGAAATCAGTACACTCCAGCTGGGCAACTA 2160

Qy        721 ProProProProSerArgTyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMet 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CCTCCTCCTCCTTCGCGTTACCCTCCAGCTTCAAACAACCCCAACTACACTAGTGGAATG 2220

Qy        741 ValHisGlyAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeu 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GTCCATGGCAACATGCAATACCAGAGCCAATCTGTTAACATGCCTCAGCTGTCTCCGTTA 2280

Qy        761 ProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProVal 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCAAATATGCCTCATAATAATTATTCCATGTACACTCAGGGTTCGTCAAATCATCCTGTT 2340

Qy        781 SerGlnProMetValGlnGlnTyrGlnProGluAlaSerMetProAsnGlnAsnTyrGly 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TCTCAGCCCATGGTCCAGCAATACCAACCAGAAGCGTCCATGCCAAACCAAAACTATGGT 2400

Qy        801 ProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsn 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 CCAATTCCAAGTTATCAGCAAGCTAATTTTCATGGCGTAACAACAAATCAGGCACAGAAC 2460

Qy        821 LeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGlu 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 TTAAACCCTTCCCAATTTCAAGCTGCCATGCAACCACCAGCAGATAAGGCAAATTTAGAG 2520

Qy        841 ProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThr 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 CCACAAAACCAAGCACTACGATTGCAGCCTATGATCTCTGGGGATGGTCAGGGTACAACA 2580

Qy        861 AspGlyGluValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsnLeu 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 GATGGGGAGGTCGATAAGAATCAGAGATACCAGTCAACACTACAATTTGCAGCAAACCTT 2640

Qy        881 LeuLeuGlnIleGlnGlnLysGlnGlnGlnGlnSerSerGlyThrProAlaGlyGlnGly 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CTTCTCCAGATACAGCAGAAACAGCAGCAACAGTCTTCAGGTACTCCGGCTGGACAGGGG 2700

Qy        901 Pro 901
              |||
Db       2701 CCT 2703
```

```
RESULT 2
US-09-920-705-1
; Sequence 1, Application US/09920705
; Publication No. US20030079252A1
; GENERAL INFORMATION:
;  APPLICANT: Amasino, Richard M.
;  APPLICANT:  Schomburg, Fritz M.
;  APPLICANT:  Michaels, Scott D.
;  APPLICANT:  Patton, David
;  TITLE OF INVENTION: Floral Induction Gene
;  FILE REFERENCE: 960296.97214
;  CURRENT APPLICATION NUMBER: US/09/920,705
;  CURRENT FILING DATE:  2001-08-02
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4593
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-920-705-1

Alignment Scores:
Pred. No.:                0              Length:          4593
Score:                    4392.00        Matches:         899
Percent Similarity:       58.72%         Conservative:    0
Best Local Similarity:    58.72%         Mismatches:      2
Query Match:              91.48%         Indels:          631
DB:                       10             Gaps:            4

US-09-920-705-3 (1-901) x US-09-920-705-1 (1-4593)

Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 60

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 120

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 180

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 240

Qy         81 SerGlnIleLysIleGluTyrAlaArgPro---------------------------- 90
              ||||||||||||||||||||||||||||
Db        241 AGTCAAATTAAGATCGAATACGCACGACCGGTTTGTTCTTATCTATATCTTCGTTTGTTC 300

Qy         90 ------------------------------------------------------------ 90

Db        301 TCTAACTTTGATTGTCTTTTGTCAACGATTATACTCTTTTTGCGAATTCATAGTCCAGGT 360

Qy         90 ------------------------------------------------------------ 90

Db        361 TCACAAAACTTTGATGATGCTTGTTTAGTCCAAAAATTTCTTGTTGAATCTGTTTTTTTT 420

Qy         90 ------------------------------------------------------------ 90

Db        421 TTCCTCATCGTACAAATCAAAGTCGAAACCTAGTTTTTTTCTATTATACGTCGTTAGCTT 480

Qy         90 ------------------------------------------------------------ 90

Db        481 AAGGCGAAACCTGATCCGATCGAAACGTCTTTTCTCAAATTACTTTGGTTATATCGAACT 540

Qy         90 ------------------------------------------------------------ 90

Db        541 CGCGCAAAGCCAAACCACAGAGAAGCTCTGCAAAATTTGATGTTAAAGCATATATAACTC 600

Qy         90 ------------------------------------------------------------ 90

Db        601 TTAGCGAATGAGCTCTGCAAAAGAATACATCAAACACATGTTTACTCTCGTTTATGCGAA 660

Qy         90 ------------------------------------------------------------ 90

Db        661 GAAGGTTTAATCCGATTGTCGTTTATCTGCAAACTTGTATCCGCGTTTAGCAGTCTGATT 720

Qy         90 ------------------------------------------------------------ 90

Db        721 TCAGGTTCGCTTCAGATGTTAAATCTCACAAGCTTGAGTATGAATGTATTGCGTACTCCA 780

Qy         90 ------------------------------------------------------------ 90
```


Db 781 CTCGAAATTCGCAATGGTGAAAATTTGGAAGCGAGCAAACTCTATCATCGGCCAAACAGA 840
Qy 90 ---------- 90
Db 841 TAAGAAATTTTGGAGTTTAAAAGTTCAGTTTCCTGCAAAAATCAAACCGCGTGGAGAAAT 900
Qy 90 ---------- 90
Db 901 TTGTCTATGGCGGTACGTAGATATCAATAACTGCATCTGCGAAACAGAAAATTCTGGCAA 960
Qy 90 ---------- 90
Db 961 CCGTATCGTCCTAACTCCTAGTATCGCTGGCACATATCCATATCAGTGTGAGGGAGTTTT 1020
Qy 90 ---------- 90
Db 1021 GGGCTTGTTTTGGCTCGATGCTTCAGGAAGACAAATTACGTGGTTGTTAAGGCGGCTAAC 1080
Qy 90 ---------- 90
Db 1081 TCTACCAATCAGAAACGCTTTATTCGAAGAACCATGTTTGTTCCTCAATTCCCATCCCTA 1140
Qy 90 ---------- 90
Db 1141 CGTACAATCTGGGCTTTCCCATTGTAGTTCCTTAGGAAGTTGACTTCTTCACAACCATTC 1200
Qy 90 ---------- 90
Db 1201 TGGGATTGGTACAGTTGCAGTGGACAAGACAAATATTCATTTGCAGGCAGACTCAAATTT 1260
Qy 90 ---------- 90
Db 1261 CAATGTCTTGCCGGCTTTGAGTACTTAAATGGAGTGTTCAGGGATTGGTTTATTTGGGAC 1320
Qy 90 ---------- 90
Db 1321 TCAGGCAATGGATAAAGACAGGAATGTTTTGCAGAAAGTATTAATGTCTTTCCGGCTTTG 1380
Qy 90 ---------- 90
Db 1381 GGGACTACATACTTCTCCTACAGACAAAATCTGATGTTTAGGCAAGAGAAACTATTAACA 1440
Qy 90 ---------- 90
Db 1441 CTGATTTAATAGAGAAAGAGGAGATGGTTTCTCCTTGCGGCAATTTTATTTGTTTAGGAA 1500
Qy 90 ---------- 90
Db 1501 AGCAATTGATATGAATTGGTGTCGTAGTGTAGTTGAAATTACTAGTTAGTTTGTGTGTTT 1560
Qy 90 ---------- 90
Db 1561 AGTTTCCTTGATGTTTGATGCTTTATTCTTGGCAACCTATCTGGGTAGTATCGCCTTCTT 1620
Qy 91 ----------AlaLysProCysLysSerLeuTrpValGlyGlyIleGly 103
Db 1621 ATCGACCTTTTCTTGTTGCAGGCAAAACCTTGTAAGAGTCTATGGGTGGGTGGAATCGGC 1680
Qy 104 ProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAsp 123
Db 1681 CCTAATGTCTCCAAGGATGACCTGGAGGAAGAGTTCAGCAAGTTTGGGAAAATCGAGGAT 1740
Qy 124 PheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAla 143
Db 1741 TTTAGGTTTCTCAGAGAACGCAAGACAGCTTTCATTGATTATTATGAGATGGATGATGCT 1800
Qy 144 LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPhe 163
Db 1801 TTACAGGCTAAGAGCATGAATGGAAAGCCTATGGGTGGTAGCTTTTTGCGTGTTGATTTT 1860
Qy 164 LeuArgSerGlnAlaProLysLys---------- 171
Db 1861 CTCCGGTCACAAGCGCCAAAAAAAGTAAGCACTCTTGTGGCATTTGATTTTTACTTTTGA 1920
Qy 171 ---------- 171
Db 1921 AAACGCTCCAGTAAACATTTTGTTTAGTTTCATAATTTGCGTCAAACTGATAGGGCTGAG 1980
Qy 171 ---------- 171
Db 1981 CTCTGTCTTGTGCCCCTAGGAGCAGTATTTACTCGTCTCTATTTCATTGTAGAGTAGGCT 2040
Qy 172 ----------GluGlnTrpAlaGly 176
Db 2041 CAACTTCTTAAGTCTGAAATCAAGTTACCTTTGTGTTATCTTCAGGAACAATGGGCTGGC 2100
Qy 177 SerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGln---------- 190
Db 2101 TCTTACGATAACAGAAATGGCAATATGAATCATAAACCGCA-GGTTAGTCTTGAATGTTG 2159
Qy 190 ---------- 190
Db 2160 AAAGTATGTCTCTTGTTACTAGTGATATGTATAGGTTACTGGTTTTGACGTTTTGTTATA 2219
Qy 191 ---------TyrProHisSerTyrGluAspPheLysGlyAspValGlnProSerLysVal 207
Db 2220 TTCTTACAGTATCCTCACTCATATGAAGACTTTAAAGGAGATGTCCAGCCAAGTAAGGTT 2279
Qy 208 LeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHisAsn 227
Db 2280 CTGTGGATTGGGTTCCCTCCTACTGCTACACAATGCAATGATGAGCAAATTCTGCACAAT 2339
Qy 228 AlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsnPheAla 247
Db 2340 GCGATGATACTCTTTGGTGAGATCGAGAGGGTAAAAAGTTACCCATCAAGGAATTTTGCA 2399
Qy 248 LeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeuGlnGlyArg 267
Db 2400 CTTGTGGAGTTTAGGAGCGCGGAGGAAGCTCGCCAATGCAAGGAAGGCCTACAGGGGAGG 2459
Qy 268 LeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProProGluGln 287
Db 2460 TTATTCAATAATCCTAGAATCAAAATTATGTACTCAAACGATGAGTTGCCTCCTGAGCAA 2519
Qy 288 AspAspThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsnAsp 307
Db 2520 GACGATACTAGTTTTTACTCTGGTATGAAACGGTCAAGGACAGATATGTTCAATAATGAT 2579
Qy 308 ProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArgProLeuArg 327
Db 2580 CCTTCATTTGTATCTTCTCCTCATTCTACTGGAATTCCTGGGTCTATGAGGCCCCTCAGA 2639
Qy 328 GlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValValGlyLysGluPro 347
Db 2640 GGTACAAATGAGCGTTCATATAATGGTGCAGAATACAATGACGTTGTTGGTAAGGAGCCA 2699
Qy 348 AsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGly 367
Db 2700 AACTGGAGGAGGCCATCTGCAAATGGAACTGGAATACTCCCATCTCCAACAGGACCTGGA 2759
Qy 368 IleLeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrp 387
Db 2760 ATCCTCCCATCTCCTGCACAAGGTACGAGGCGCCCTATGAGGTCAAACCCCGATTCTTGG 2819
Qy 388 GluGlyTyrAspProAlaGlnLeuValArgGluSerLysArgThrArgArgAspGlySer 407
Db 2820 GAAGGATATGATCCTGCTCAGTTGGTCAGAGAAAGTAAACGAACCAGAAGAGATGGATCA 2879
Qy 408 ValAspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGlySerValAla 427
Db 2880 GTGGACGGTTTTACTCCAATGGGTGTCGATGAGAGGTCATTTGGTCGAGGTTCAGTTGCT 2939
Qy 428 AlaArgProIleArgGlyProProAspSerAspHisIleTrpArgGlyMetIleAlaLys 447
Db 2940 GCTAGACCTATCCGTGGCCCCCCTGATTCTGATCACATATGGAGAGGAATGATTGCCAAG 2999

```
Qy       448 GlyGlyThrProValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLys 467
Db      3000 GGTGGAACTCCCGTCTGTTGTGCTCGTTGTGTACCTATGGGAAAGGGGATTGAAACTAAA 3059

Qy       468 Leu------------------------------------------------------------ 468
Db      3060 CTGTGAGTACTAATTTCTAGCACTTTAACCCTTCTAGTGTTTTCTTTTTCAGAGCGATTT 3119

Qy       468 ------------------------------------------------------------ 468
Db      3120 ATATATTTTCCATTTCATTCTCGATGGAAGTAACATTATTATAGATAGTACATTTTTATT 3179

Qy       468 ------------------------------------------------------------ 468
Db      3180 TTACTATTACTTGTTTAGTTTCTGAGATGTCTTGATTTTCATGGTGTTGATTCATTTTTG 3239

Qy       469 ----------------------------------------------------------ProGluVa 471
Db      3240 GCATTGCCCTCAATTACTGACTTTGTTTTTTTTTTAATAATTGATTTATAGGCCTGAGGT 3299

Qy       471 lValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIl 491
Db      3300 CGTCAATTGTTCAGCAAGAACTGATTTGAATATGCTCGCTAAACATTACGCCGTTGCCAT 3359

Qy       491 eGlyCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGl 511
Db      3360 TGGATGTGAGATCGTTTTTTTCGTACCAGACAGGGAAGAAGATTTTGCGTCTTACACTGA 3419

Qy       511 uPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyTh 531
Db      3420 ATTTCTCCGGTACCTTAGCTCAAAAGATCGGGCGGGTGTTGCCAAATTAGATGATGGTAC 3479

Qy       531 rThrLeuPheLeuValProProSerAspPheLeuThrAspValLeuGlnValThrArgGl 551
Db      3480 AACTTTATTCTTGGTGCCTCCATCAGATTTCTTAACTGATGTACTCCAAGTGACCCGTCA 3539

Qy       551 nGluArgLeuTyrGlyValValLeuLysLeuProProProAlaValProValThrAlaSe 571
Db      3540 AGAACGGCTATATGGTGTTGTTCTCAAGTTACCCCCGCCAGCCGTTCCTGTTACAGCATC 3599

Qy       571 rTyrArgGlnGluSerGlnSerAsnProLeuHisTyrMetAspGlnAlaArgAspSerPr 591
Db      3600 ATACAGACAAGAATCTCAGTCCAATCCTCTGCATTATATGGATCAAGCCCGGGATTCACC 3659

Qy       591 oAlaAsnAlaSerHisSerLeuTyrProProArgGluAsnTyrIleArgGlyAlaProGl 611
Db      3660 TGCCAATGCTAGTCACAGTTTATATCCTCCTAGGGAAAATTACATTAGGGGTGCACCAGA 3719

Qy       611 uHisLeuThrAlaAlaSerLysProSerValSerGluProLeuArgIleProAsnAsnAl 631
Db      3720 ACATTTGACAGCTGCTTCAAAACCATCTGTTAGCGAGCCTCTCAGAATACCTAATAATGC 3779

Qy       631 aAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAlaThrLeuAlaSerIleLe 651
Db      3780 AGCGCCTCAAGCTGGGGTTAGTTTAACTCCGGAGCTTTTAGCCACTCTGGCATCTATTCT 3839

Qy       651 uProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMetSerGlyProSerTh 671
Db      3840 CCCTGCAACTTCTCAACCTGCTGCCCCTGAGAGTCACCAACCTATGTCAGGACCTTCAAC 3899

Qy       671 rValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyGluAlaProSerGlnAl 691
Db      3900 AGTTGTTTCCACAGCACATCAGTCCAATGGACTGTACAATGGAGAAGCACCGTCTCAAGC 3959

Qy       691 aTrpLysArgGlyProGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTyrGl 711
Db      3960 TTGGAAAAGAGGTCCACAAACAGTTCATGATGCGTCAAATCAGTCATTCCAACAATACGG 4019

Qy       711 yAsnGlnTyrThrProAlaGlyGlnLeuProProProProSerArgTyrProProAlaSe 731
Db      4020 AAATCAGTACACTCCAGCTGGGCAACTACCTCCTCCTCCTTCGCGTTACCCTCCAGCTTC 4079

Qy       731 rAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMetGlnTyrGlnSerGlnSe 751
Db      4080 AAACAACCCCAACTACACTAGTGGAATGGTCCATGGCAACATGCAATACCAGAGCCAATC 4139

Qy       751 rValAsnMetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyrSerMetTy 771
Db      4140 TGTTAACATGCCTCAGCTGTCTCCGTTACCAAATATGCCTCATAATAATTATTCCATGTA 4199

Qy       771 rThrGlnGlySerSerAsnHisProValSerGlnProMetValGlnGlnTyrGlnProGl 791
Db      4200 CACTCAGGGTTCGTCAAATCATCCTGTTTCTCAGCCCATGGTCCAGCAATACCAACCAGA 4259

Qy       791 uAlaSerMetProAsnGlnAsnTyrGlyProIleProSerTyrGlnGlnAlaAsnPheHi 811
Db      4260 AGCGTCCATGCCAAACCAAAACTATGGTCCAATTCCAAGTTATCAGCAAGCTAATTTTCA 4319

Qy       811 sGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGl 831
Db      4320 TGGCGTAACAACAAATCAGGCACAGAACTTAAACCCTTCCCAATTTCAAGCTGCCATGCA 4379

Qy       831 nProProAlaAspLysAlaAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMe 851
Db      4380 ACCACCAGCAGATAAGGCAAATTTAGAGCCACAAAACCAAGCACTACGATTGCAGCCTAT 4439

Qy       851 tIleSerGlyAspGlyGlnGlyThrThrAspGlyGluValAspLysAsnGlnArgTyrGl 871
Db      4440 GATCTCTGGGGATGGTCAGGGTACAACAGATGGGGAGGTCGATAAGAATCAGAGATACCA 4499

Qy       871 nSerThrLeuGlnPheAlaAlaAsnLeuLeuLeuGlnIleGlnGlnLysGlnGlnGlnGl 891
Db      4500 GTCAACACTACAATTTGCAGCAAACCTTCTTCTCCAGATACAGCAGAAACAGCAGCAACA 4559

Qy       891 nSerSerGlyThrProAlaGlyGlnGlyPro 901
Db      4560 GTCTTCAGGTACTCCGGCTGGACAGGGGCCT 4590
```

RESULT 3
US-09-920-705-4/c
; Sequence 4, Application US/09920705
; Publication No. US20030079252A1
; GENERAL INFORMATION:
; APPLICANT: Amasino, Richard M.
; APPLICANT: Schomburg, Fritz M.
; APPLICANT: Michaels, Scott D.
; APPLICANT: Patton, David
; TITLE OF INVENTION: Floral Induction Gene
; FILE REFERENCE: 960296.97214
; CURRENT APPLICATION NUMBER: US/09/920,705
; CURRENT FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: FPA antisense fragment
US-09-920-705-4

Alignment Scores:

Pred. No.:	6.03e-121	Length:	801
Score:	1436.00	Matches:	265
Percent Similarity:	99.62%	Conservative:	0
Best Local Similarity:	99.62%	Mismatches:	1
Query Match:	29.91%	Indels:	0
DB:	10	Gaps:	0

US-09-920-705-3 (1-901) x US-09-920-705-4 (1-801)

```
Qy       197 AspPheLysGlyAspValGlnProSerLysValLeuTrpIleGlyPheProProThrAla 216
Db       799 GACTTTAAAGGAGATGTCCAGCCAAGTAAGGTTCTGTGGATTGGGTTCCCTCCTACTGCT 740

Qy       217 ThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGlu 236
```

```
Db        739 ACACAATGCAATGATGAGCAAATTCTGCACAATGCGATGATACTCTTTGGTGAGATCGAG 680

Qy        237 ArgValLysSerTyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGlu 256

Db        679 AGGGTAAAAAGTTACCCATCAAGGAATTTTGCACTTGTGGAGTTTAGGAGCGCGGAGGAA 620

Qy        257 AlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIle 276

Db        619 GCTCGCCAATGCAAGGAAGGCCTACAGGGGAGGTTATTCAATAATCCTAGAATCAAAATT 560

Qy        277 MetTyrSerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMet 296

Db        559 ATGTACTCAAACGATGAGTTGCCTCCTGAGCAAGACGATACTAGTTTTTACTCTGGTATG 500

Qy        297 LysArgSerArgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSer 316

Db        499 AAACGGTCAAGGACAGATATGTTCAATAATGATCCTTCATTTGTATCTTCTCCTCATTCT 440

Qy        317 ThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGly 336

Db        439 ACTGGAATTCCTGGGTCTATGAGGCCCCTCAGAGGTACAAATGAGCGTTCATATAATGGT 380

Qy        337 AlaGluTyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGly 356

Db        379 GCAGAATACAATGACGTTGTTGGTAAGGAGCCAAACTGGAGGAGGCCATCTGCAAATGGA 320

Qy        357 ThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThr 376

Db        319 ACTGGAATACTCCCATCTCCAACAGGACCTGGAATCCTCCCATCTCCTGCACAAGGTACG 260

Qy        377 ArgArgProMetArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuVal 396

Db        259 AGGCGCCCTATGAGGTCAAACCCCGATTCTTGGGAAGGATATGATCCTGCTCAGTTGGTC 200

Qy        397 ArgGluSerLysArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyVal 416

Db        199 AGAGAAAGTAAACGAACCAGAAGAGATGGATCAGTGGACGGTTTTACTCCAATGGGTGTC 140

Qy        417 AspGluArgSerPheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAsp 436

Db        139 GATGAGAGGTCATTTGGTCGAGGTTCAGTTGCTGCTAGACCTATCCGTGGCCCCCCTGAT 80

Qy        437 SerAspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrProValCysCysAlaArg 456

Db         79 TCTGATCACATATGGAGAGGAATGATTGCCAAGGGTGGAACTCCCGTCTGTTGTGCTCGT 20

Qy        457 CysValProMetGlyLys 462

Db         19 TGTGTACCTATGGGAAAG 2
```

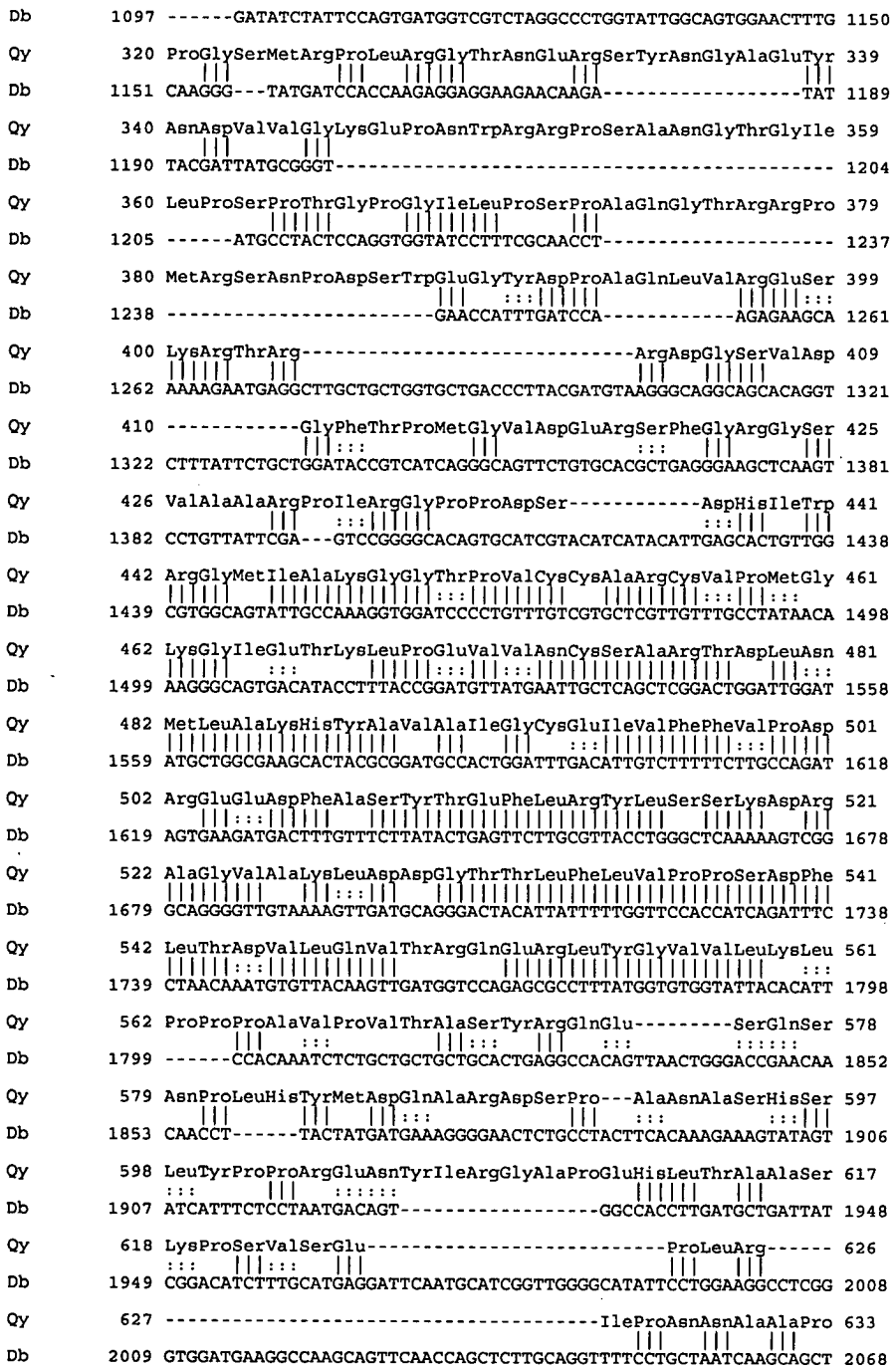
RESULT 4
US-10-425-115-123950
; Sequence 123950, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 123950
; LENGTH: 3463
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_44523C.1
US-10-425-115-123950

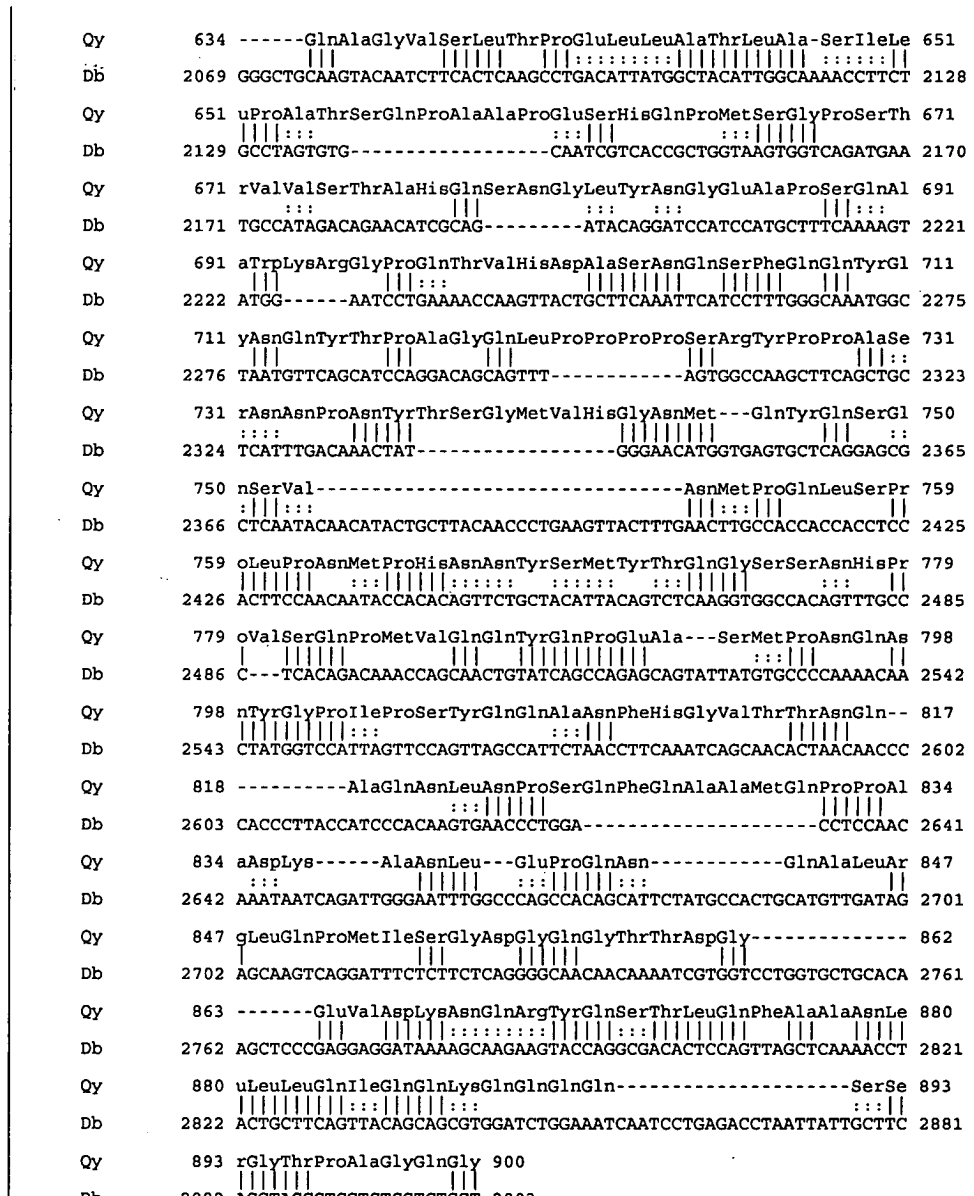
Alignment Scores:

Pred. No.:	1.66e-106	Length:	3463
Score:	1286.50	Matches:	352
Percent Similarity:	49.80%	Conservative:	140
Best Local Similarity:	35.63%	Mismatches:	298
Query Match:	26.80%	Indels:	198
DB:	18	Gaps:	40

US-09-920-705-3 (1-901) x US-10-425-115-123950 (1-3463)

```
Qy         16 GlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThrGluSerAspLeuThr 35

Db        227 GAAACGAACACGCTCTGGGTGGGCAACCTGCCCTCGCACGTGACCGAGGGCGATTTGTTG 286

Qy         36 GluLeuPheGlyArgTyrGlyAspIleAsp---ArgIleThrValTyrSerSerArgGly 54

Db        287 GCGCTGTTCGGGCCGCACGGCGCGCTCGATTGCGCCCTTGCACGAGCCGGATCTCGCAGC 346

Qy         55 PheAlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGln 74

Db        347 TACGCGTTCGTCCTCTTCCGTTCCCCAGCCGAGGCTCGCGCAGCCGTCGAGGCCACCCGG 406

Qy         75 GlyAlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAlaArgProAlaLysProCys 94

Db        407 GGTGAAAAGGTCAAGGGGGCCGCCATGCGCACCGAGTTCGCGCGGCCGGCTAGAGCTGTT 466

Qy         95 LysSerLeuTrpValGlyGlyIleGlyProAsnValSerLysAspAspLeuGluGluGlu 114

Db        467 AGAAATCTGTGGGTTGGTGGCATTAGCCCATCCATTTCAAAGGAGGAGCTTGAGGGGGAG 526

Qy        115 PheSerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLysThrAlaPhe 134

Db        527 TTTCTAAAGTTCGGAAAAGTTGAAGGTGTTGCGTTCTCCCAAGATCAGACTTCAGCATAC 586

Qy        135 IleAspTyrTyrGluMetAspAspAlaLeuGlnAla---LysSerMetAsnGlyLysPro 153

Db        587 ATAGATTTTGAGAAGCTTGAAGACGCAATTTCTGCTCACAGATCCTTGAATGGAAAAATG 646

Qy        154 MetGlyGlySerPheLeuArgValAspPheLeuArgSerGlnAlaProLysLysGluGln 173

Db        647 CTAGGTGGCAAGGAATTGTGTGTTGATTTCCAGAGGTCCAAGGGAAGAGCGGAATGGTCA 706

Qy        174 TrpAlaGlySerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGlnTyrProHis 193

Db        707 GAAGCTAGCAGCTTCAATGGTAGA----------GTACCAGGG 739

Qy        194 SerTyrGluAspPheLysGly---------- 200

Db        740 CCAGTGGGTGATAAGAGAGGCAGTGGCCCTCTGAAGAGTTCTGCTGGAGTGCGGATGCGA 799

Qy        201 AspValGlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220

Db        800 GAGGCACAACCTACCAATGTTCTCTGGGTTGGTTTT---CCAGGTTCTTATAGAGTCATT 856

Qy        221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240

Db        857 GATGAGGAGGCACTTAAGCATGCCATGTCAGTATTCGGTGTCGTTACGAAGATAAAAGTT 916

Qy        241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260

Db        917 TTCCAGACAAGGCAGTATGCTTTTGTTGAGTTTGCAAATGTTGCGGAAGCTTGTAATGCT 976

Qy        261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280

Db        977 AAGATGAATCTAGATGGTCATCTTTTCAATGATCCCAGGATTCAGATTCTTTTCTCAAAC 1036

Qy        281 AspGluLeuProProGluGln---AspAspThrSerPheTyrSerGlyMetLysArgSer 299

Db       1037 AGTGGGCTTGCACCAAACAAATTGGACAACCCAACATCGGTTGCTGGATTTCCTAGTTCA 1096

Qy        300 ArgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIle 319
```

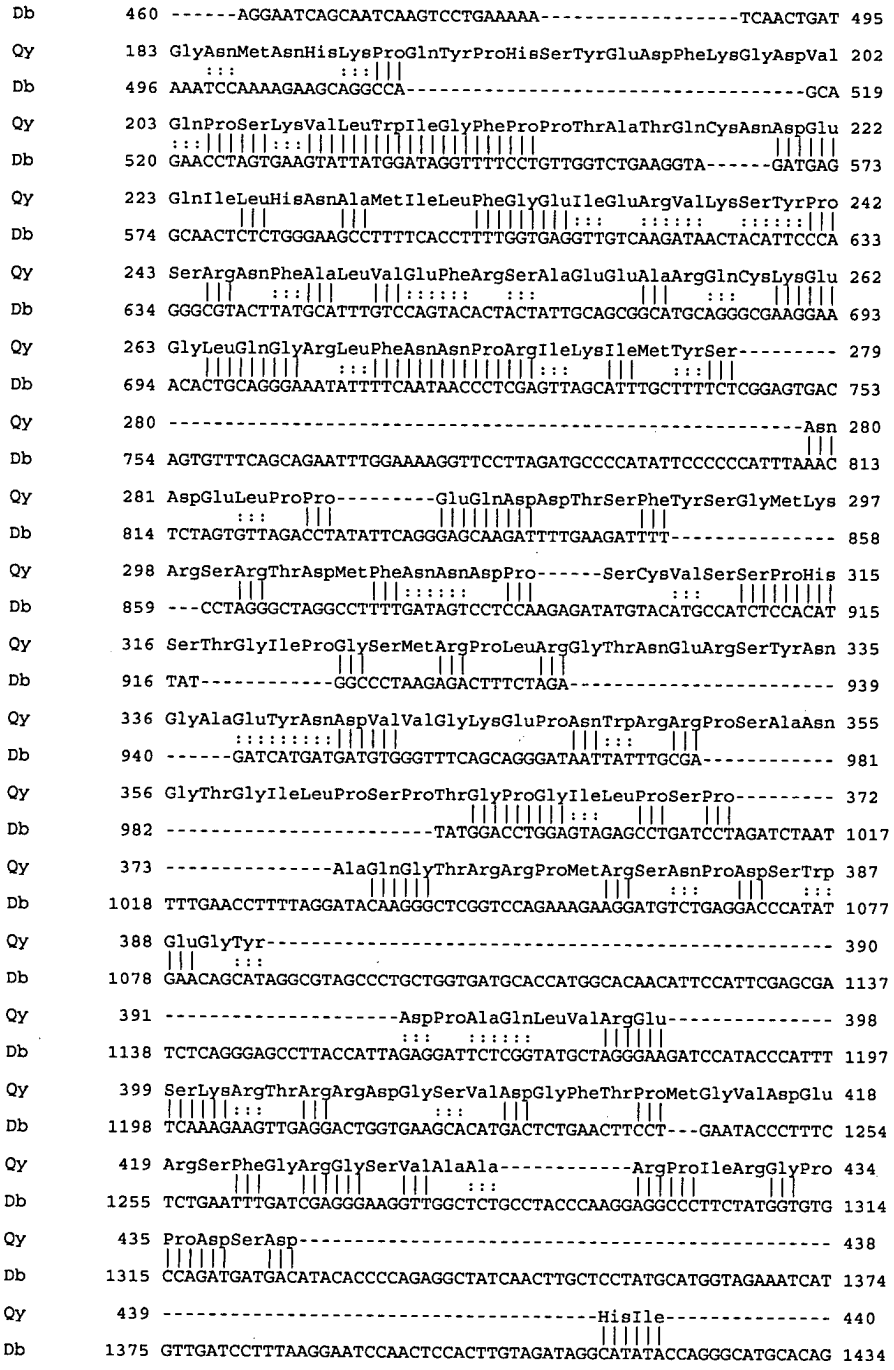




RESULT 5
US-10-437-963-36321
; Sequence 36321, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua

```
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 36321
;   LENGTH: 3084
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_40159C.1
US-10-437-963-36321

Alignment Scores:
Pred. No.:              4.06e-106       Length:         3084
Score:                  1281.50         Matches:        339
Percent Similarity:     51.47%          Conservative:   150
Best Local Similarity:  35.68%          Mismatches:     306
Query Match:            26.69%          Indels:         155
DB:                     18              Gaps:           38

US-09-920-705-3 (1-901) x US-10-437-963-36321 (1-3084)

Qy         16 GlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThrGluSerAspLeuThr 35
Db        178 GAGACGAACACGCTGTGGGTGGGCAACCTGCCCGCGCAGGCGGCGGAGGACGACGTGATG 237

Qy         36 GluLeuPheGlyArgTyrGlyAspIleAspArgIleThrValTyrSerSer---ArgGly 54
Db        238 GCGTCCTTCTCCCCGCACGGCGCGCTCGACTGCGTCATGGCCCGCGCGGGCCCGCGGAGC 297

Qy         55 PheAlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGln 74
Db        298 TACGCGTTCGTGCTGTTCCGCTCCGTCCCGGAGGCCCGCGCCGCCCTCGACGCGCTCCAG 357

Qy         75 GlyAlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAlaArgProAlaLysProCys 94
Db        358 GGCTCCAAGGTCAAGGGCAGCGTCGTCCGGCTCGAGTTCGCCCGGCCGGCTAGAGCTGTT 417

Qy         95 LysSerLeuTrpValGlyGlyIleGlyProAsnValSerLysAspAspLeuGluGluGlu 114
Db        418 AAGAACTTATGGGTTGGTGGCATTAGCTCCTCAATTTCAAAGGAGGAGCTGGAGGAGGAG 477

Qy        115 PheSerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLysThrAlaPhe 134
Db        478 TTTAAGAAATTTGGAAAGGTTGATGGTATTGCATTCTCTCGTGATCAGACTTCCGCGTAC 537

Qy        135 IleAspTyrTyrGluMetAspAspAlaLeuGlnAla---LysSerMetAsnGlyLysPro 153
Db        538 ATTGATTTTGACAAGCTGGAAGATGCAATTTCGGCACATAGAGCCTTGAATGGAAGAGTT 597

Qy        154 MetGlyGlySerPheLeuArgValAspPheLeuArgSerGlnAlaProLysLysGluGln 173
Db        598 TTAGGTGGACAAGAACTGTGTGTTGATTTCCAAAGGTCCAGGGGC------AGAGCGGAA 651

Qy        174 Trp------AlaGlySerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGlnTyr 191
Db        652 TGGTTGGAGACAGGCAGCTTCAATGGTAGAACTGGACCTGCAAAG-------------- 696

Qy        192 ProHisSerTyrGluAspPheLysGlyAspValGlnProSerLysValLeuTrpIleGly 211
Db        697 ------GGTTATGGAGTACGAAATCGAGAATCCAATCCCACCAATGTTCTTTGGGTAGGT 750

Qy        212 PheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeu 231
Db        751 TTCCCTAATACAGCTAAA------ATCAATGAGGAGGCACTTCGGCAAGCGATGGCTGTA 804

Qy        232 PheGlyGluIleGluArgValLysSerTyrProSerArgAsnPheAlaLeuValGluPhe 251
Db        805 CATGGTGCTGTTACAAATACAAAAGTTTCCCAACAAGGCAATATGCTTTTGTGGAGTTT 864

Qy        252 ArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsn 271
Db        865 GCAACAGTTGGGGAAGCTTCTAATGCTAAGAAAAATCTAGATGGTCGTCTTTTCAATGAT 924

Qy        272 ProArgIleLysIleMetTyrSerAsnAspGluLeuProProGluGln---AspAspThr 290
Db        925 CAGAGAATTCAAATTCTTTTCTCGAACAGTGAGCTTGCGCCAAATAAATTAGATAATCCA 984

Qy        291 SerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsnAspProSerCys 310
Db        985 ACAGCAGTTTCTGGATTTCCAAAATCA------GAGATGTATTATGATGATGGTCAGTAT 1038

Qy        311 ValSerSerProHisSerThrGlyIleProGlySerMetArgProLeuArgGlyThrAsn 330
Db       1039 GGCGCTTCTGATTAT--------------------TTTGATCCTCGACGGGGAAGATCA 1077

Qy        331 GluArgSerTyrAsnGlyAlaGluTyrAsnAspValValGlyLysGluProAsnTrpArg 350
Db       1078 AGATAC------------TTTGAGTACAGTGGTGTC---------------------- 1101

Qy        351 ArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuPro 370
Db       1102 ----------------------------------CCTGTTTCTGGTGGTATTCTTCCA 1125

Qy        371 SerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrpGluGlyTyr 390
Db       1126 TCACCGGAATCTGGA------AATCCTCTCTTAACTGGACGATCTGCTCAGAGTACATTC 1179

Qy        391 AspProAlaGlnLeuValArgGluSerLysArgThrArgArgAspGlySerValAsp--- 409
Db       1180 GATCCA------------AGAGAAGCCAAGAGGTTGCGGTTGGATGCTGCAGCTGACCCT 1227

Qy        410 ----------------------GlyPheThrProMetGlyValAspGluArgSerPhe 421
Db       1228 TATGATACAAGGGCAGGTAGTGAGGGTCTTTATTCTGCTGGGTACTCTCAGCGCGAGAGT 1287

Qy        422 GlyArgGlySerValAlaAlaArgPro---------IleArgGlyProProAspSer--- 437
Db       1288 GCACGTTCTGAGAGGAGCTCAAGTCCTGCTATTCGAATCCATGGCACAGTGCACCGAACA 1347

Qy        438 ---------AspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrProValCysCys 454
Db       1348 TCATACCTTGAGCACTTCTGGCGTGGTAGCATTGCCAAAGGTGGATCTCCTCCTCTGGAT 1407

Qy        455 AlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluValValAsnCys 474
Db       1408 CTTACACATATTCCACTG--------------TTTTGCAGACCGGATGTTGTTAATTGT 1452

Qy        475 SerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGlu 494
Db       1453 TCCGCTAGAACTGGACTAGATATGCTGGCAAAGCATTATAGAGACGCTTCAGGGTTTGAT 1512

Qy        495 IleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArg 514
Db       1513 ATCGTCTTCTTCTTGCCAGATAGTGAAGATGACTTTGTTTCTTACACTGAATTTTTGCGC 1572

Qy        515 TyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPhe 534
Db       1573 TACTTAGGCTCAAAAAGTCGGGCAGGGGTTGTAAAGGTTGATGGGGGGACTACTTTATTT 1632

Qy        535 LeuValProProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeu 554
Db       1633 TTGGTCCCACCTTCCGATTTTTTACGAAATGTTTTGCAAGTTGATGGTCCAGAACGTCTT 1692

Qy        555 TyrGlyValValLeuLysLeuPro-----------------ProProAlaValProVal 568
Db       1693 TACGGTGTAGTATTACACATTCCGCAAATGTCTGCTGCTGCTCCTGCTTCTGCTCCCACT 1752

Qy        569 ThrAlaSerTyrArgGlnGlu---------SerGlnSerAsnProLeuHisTyrMetAsp 585
```

```
Db       1753 CCTGCTGTACAAAGGCCACAACTAACTGCACCAGAATCACAACCT------TACTACGAT 1806

Qy        586 GlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGluAsnTyr 605

Db       1807 GAA---AGGGAAATTCCTTTGCAGAGAAGGTATAGTATGATTACGCCTAGCAACAATCAC 1863

Qy        606 IleArgGlyAlaProGluHisLeuThrAla---------------------------- 615

Db       1864 CATCGAGATGCT---GATCACCGTGGATCTTTACGTGAGGATTCATTGCATCAATTGGGG 1920

Qy        616 ------AlaSerLysProSerValSerGluPro--------------------LeuArg 626

Db       1921 CAAATTCTTGCACGACCTCGGGTGGATGAAGGACAAGTAGTTCAGCCAAACCTTGCCGGT 1980

Qy        627 IleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAlaThr 646

Db       1981 ATCCCTACAAACGCAGGACTACAGGTACAGCCTTCACTCCAGCCTGATATGATAGCCACT 2040

Qy        647 LeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMet 666

Db       2041 TTAGCTAAACTTTGCCAAGTGGGCAGTCATCAGCTCTGGTCACTGGCCAGCTTCCCCTG 2100

Qy        667 SerGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyGlu 686

Db       2101 AGC------TCAACAGACCAACCTGCATTAACCCAAATGAAT---------GATGCATCT 2145

Qy        687 AlaProSerGlnAlaTrpLysArgGlyProGlnThrValHisAspAlaSerAsnGlnSer 706

Db       2146 ACTCTTGCAAAAGTATGGAGA------CCTGAAAATCAAGCAATGGCTTCTACTTCATCC 2199

Qy        707 PheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnLeuProProProProSerArg 726

Db       2200 TTAGAGCAGATCGGTAAT---TTCCAGCATTCAGGACAA---------CAGTTCAGCAAA 2247

Qy        727 TyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMetGln 746

Db       2248 CAAGCTGGAGCAGTTCATCTGCCAAACTAT------GGGAACTTGGCAGGTGCACAGGAG 2301

Qy        747 TyrGlnSerGln-----------------------SerValAsnMetProGlnLeuSer 758

Db       2302 CACCCAACACAACATAGTGCTTACAATCCTGAAATGACTCTAAACTTGCCACCACCACCA 2361

Qy        759 ProLeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHis 778

Db       2362 CCACCTCCTACGCTGCCACCCAGTTCTGCTATATTATCATCGCAAGTTGGACACAGTTTG 2421

Qy        779 ProValSerGlnProMetValGlnGlnTyrGlnProGluAla---SerMetProAsnGln 797

Db       2422 CCT---ACGCAGATGAGTCAGCAACAGTATCAACCAGAGCAGTATTATATGACTCAAAGC 2478

Qy        798 AsnTyrGlyProIleProSerTyrGlnGlnAlaAsnPheHisGly-------------- 812

Db       2479 AATTATGGTCAATTGGCCACAGTTAGCAGTTCTAATCTTCAGGCGCACCATCAACAAATT 2538

Qy        813 ValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGlnPro 832

Db       2539 GTTGCGACTCCTGCTGCACAAGCTCCTGTTGCTGCGCAATTCCCTCCTGCTATGCAAGCT 2598

Qy        833 ProAlaAspLysAlaAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMetIle 852

Db       2599 CCTGCT-----------------------------------GCCGCACAAGCTCCTGTT 2622

Qy        853 SerGlyAspGlyGlnGlyThrThrAspGlyGluValAspLysAsnGlnArgTyrGlnSer 872

Db       2623 GCT------GCACAAGCTTCTGCTGAC---GAGGCAGAACGGAACAGGAAGTATCAGGCG 2673

Qy        873 ThrLeuGlnPheAlaAlaAsnLeuLeuGlnIleGlnGlnLysGlnGlnGlnGlnSer 892

Db       2674 ACTCTTCAGTTGGCTCAGCGCCTGTTGGGTCAGCTACAACAGAAGCCTGGAAATCAACCT 2733

Qy        893 SerGly------------ThrProAlaGly 898

Db       2734 TAAGGCTATCTTGGAAATCAACCTTCAGGC 2763

RESULT 6
US-10-437-963-79215
; Sequence 79215, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 79215
;   LENGTH: 3315
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_78947C.1
US-10-437-963-79215

Alignment Scores:
Pred. No.:                1.9e-62        Length:          3315
Score:                    802.00         Matches:         298
Percent Similarity:       37.29%         Conservative:    120
Best Local Similarity:    26.58%         Mismatches:      325
Query Match:              16.70%         Indels:          378
DB:                       18             Gaps:            45

US-09-920-705-3 (1-901) x US-10-437-963-79215 (1-3315)

Qy          7 ProPheArgAlaAspAspSerGlyPhe-----------GlnSerAsnAsnLeuTrpVal 22

Db        121 CCGCCGCGGCGCTCCTCCTCCGGGTGGGGGGTGGCACCGCCGTCGCGGCACCTGTGGGTG 180

Qy         23 GlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArgTyrGly 42

Db        181 GGCAGCCTCTCCCCGGGCGTCGCCGCGGCCGACCTCTCGGAGCTCTTCCTCCGGTGCGGC 240

Qy         43 AspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyrArgHis 62

Db        241 GACGTCGAGGGCATCTCCCGTGACCCCGGCCGGAGCTTCGCGTTCGTGACGTTCGCGCGG 300

Qy         63 ValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGln 82

Db        301 GAGGAGGACGCCGTGGCGGCGGTGCGGGAGCTGCAGGGGATCCACCTCCGCGGGGCGCCC 360

Qy         83 IleLysIleGluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGlyGlyIle 102

Db        361 ATTAGGATCGAGTTTTCCAAG----------------------------GGGGATAAA 390

Qy        103 GlyProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGlu 122

Db        391 GGTTCAAGTAGCTCTATGGATGAC-----------------AGATACTCACAACATGCT 432

Qy        123 AspPheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAsp 142

Db        433 GATCAAAGACGTTTTACTGAACGAGGA---------------------------------- 459

Qy        143 AlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAsp 162

Db        459 ------------------------------------------------------------ 459

Qy        163 PheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsn 182
```


Db 460 ------AGGAATCAGCAATCAAGTCCTGAAAAA----------TCAACTGAT 495
Qy 183 GlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
Db 496 AAATCCAAAAGAAGCAGGCCA----------GCA 519
Qy 203 GlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGlu 222
Db 520 GAACCTAGTGAAGTATTATGGATAGGTTTTCCTGTTGGTCTGAAGGTA------GATGAG 573
Qy 223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrPro 242
Db 574 GCAACTCTCTGGGAAGCCTTTTCACCTTTTGGTGAGGTTGTCAAGATAACTACATTCCCA 633
Qy 243 SerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGlu 262
Db 634 GGGCGTACTTATGCATTTGTCCAGTACACTACTATTGCAGCGGCATGCAGGGCGAAGGAA 693
Qy 263 GlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSer--------- 279
Db 694 ACACTGCAGGGAAATATTTTCAATAACCCTCGAGTTAGCATTTGCTTTTCTCGGAGTGAC 753
Qy 280 ----------Asn 280
Db 754 AGTGTTTCAGCAGAATTTGGAAAAGGTTCCTTAGATGCCCCATATTCCCCCCATTTAAAC 813
Qy 281 AspGluLeuProPro---------GluGlnAspAspThrSerPheTyrSerGlyMetLys 297
Db 814 TCTAGTGTTAGACCTATATTCAGGGAGCAAGATTTTGAAGATTTT---------- 858
Qy 298 ArgSerArgThrAspMetPheAsnAsnAspPro------SerCysValSerSerProHis 315
Db 859 ---CCTAGGGCTAGGCCTTTTGATAGTCCTCCAAGAGATATGTACATGCCATCTCCACAT 915
Qy 316 SerThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsn 335
Db 916 TAT----------GGCCCTAAGAGACTTTCTAGA---------- 939
Qy 336 GlyAlaGluTyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsn 355
Db 940 ------GATCATGATGATGTGGGTTTCAGCAGGGATAATTATTTGCGA---------- 981
Qy 356 GlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerPro--------- 372
Db 982 ----------TATGGACCTGGAGTAGAGCCTGATCCTAGATCTAAT 1017
Qy 373 ----------AlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrp 387
Db 1018 TTTGAACCTTTTAGGATACAAGGGCTCGGTCCAGAAAGAAGGATGTCTGAGGACCCATAT 1077
Qy 388 GluGlyTyr---------- 390
Db 1078 GAACAGCATAGGCGTAGCCCTGCTGGTGATGCACCATGGCACAACATTCCATTCGAGCGA 1137
Qy 391 ----------AspProAlaGlnLeuValArgGlu---------- 398
Db 1138 TCTCAGGGAGCCTTACCATTAGAGGATTCTCGGTATGCTAGGGAAGATCCATACCCATTT 1197
Qy 399 SerLysArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGlu 418
Db 1198 TCAAAGAAGTTGAGGACTGGTGAAGCACATGACTCTGAACTTCCT---GAATACCCTTTC 1254
Qy 419 ArgSerPheGlyArgGlySerValAlaAla----------ArgProIleArgGlyPro 434
Db 1255 TCTGAATTTGATCGAGGGAAGGTTGGCTCTGCCTACCCAAGGAGGCCCTTCTATGGTGTG 1314
Qy 435 ProAspSerAsp---------- 438
Db 1315 CCAGATGATGACATACACCCCAGAGGCTATCAACTTGCTCCTATGCATGGTAGAAATCAT 1374
Qy 439 ----------HisIle---------- 440
Db 1375 GTTGATCCTTTAAGGAATCCAACTCCACTTGTAGATAGGCATATACCAGGGCATGCACAG 1434
Qy 440 ---------- 440
Db 1435 GACAGCTTTTCTAGGCATGTAGAAGTGGAAAGATCAACTCCTGAATACCATGAACCCCTT 1494
Qy 441 ----------TrpArgGlyMetIleAlaLysGlyGlyThrProValCysCys 454
Db 1495 CTCAAGGAAGAATGGAAATGGGATGGTACAATAGCAAAGGGAGGCACACCAATTTGCCGA 1554
Qy 455 AlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluValValAsnCys 474
Db 1555 GCGCGATGCTTCCCTGTTGGGAAGGTTCTTAACTTCATGCTGCCCGAATTTTTGGATTGC 1614
Qy 475 SerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGlu 494
Db 1615 ACTGCTAGGACAAGCCTGGAGATGCTCTCTAAGCACTATTACCAAGCTGCCAGCAGCTGG 1674
Qy 495 IleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArg 514
Db 1675 GTGGTGTTTTTTGTTCCAGAAAATGATGCTGACATGGCAGCCTATAATGAATTCATGAAT 1734
Qy 515 TyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPhe 534
Db 1735 TACCTTGGTGATAAGCAGCGTGCAGCAGTTTGTAAACTTGGAGAAAGGAGCAGCTTATTT 1794
Qy 535 LeuValProProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeu 554
Db 1795 CTTGTTCCACCCTCAGACTTCTCTGAACAAGTACTGAGGGTTCCAGGTAAAGTCAGCATA 1854
Qy 555 TyrGlyValValLeuLysLeuProProProAlaValProValThrAlaSerTyrArgGln 574
Db 1855 TCTGGAGTCATTCTGAAGTTTGAGCAGTCAGATCCAGAAGTTTCCTCGCCAACTCGCAAA 1914
Qy 575 GluSer---------- 576
Db 1915 CCAGAAACATTTGTGAGTCATTTGAACCATGATGTTCGTGCTCATGAGGATCTAGATGCA 1974
Qy 577 ------GlnSerAsnPro------LeuHisTyrMetAspGlnAlaArgAsp--------- 589
Db 1975 TTGAGAAGAATCAACCCACCAGATATCAGGCCACTTCCTCAGGGTTCAGATTATCTCGGG 2034
Qy 590 ---SerProAlaAsn---------AlaSerHisSerLeuTyrProProArgGluAsnTyr 605
Db 2035 TTGTCGCCTGGAAGCTATAATCCAGCAAGTGCACATTTGGTTCCGCCTTACAAG------ 2088
Qy 606 IleArgGlyAlaProGluHisLeuThrAla---------- 615
Db 2089 TTTGGAAATGCTCCTTCATATCTAGAATCTGAATTAGCTCATCAAAAGCATCCACCTGAC 2148
Qy 616 ----------AlaSerLysProSerValSerGluProLeuArgIleProAsn--- 629
Db 2149 TCCCACAGGGAGATAGCACATGACAAGCAGCAGCAACACCCAGATGTATTGCCCTCAAGA 2208
Qy 630 ----------AsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeu 643
Db 2209 TGGTCAGATAACATTTACAATCCAAGTCCAGGTTCTGGA----------AATTTG 2253
Qy 644 LeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHis 663
Db 2254 AATTATTTGGCTGAGAGTGCGATCCCACATACATCAACTGATAGGACACCAGAGGCATAC 2313
Qy 664 Gln------ProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsnGly 681
Db 2314 TCATTTGCTCCTCAAGGAGTACCAAAA------GTGAGTACA---------- 2349
Qy 682 LeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHisAsp 701
Db 2350 ------TCAGGGTATGCACCA----------GTTGCAGATGAG 2376
Qy 702 AlaSerAsnGlnSerPhe------GlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGln 719
Db 2377 GCATCAAACATGTCCTACCCTCCCATGCAACCTGCATCACAGCAGGTAGTTAGACCTCAA 2436

```
Qy        720 LeuProProPro---ProSerArgTyrProPro--------------------------- 729
              ||||||      |||         ||||||
Db       2437 CAACCTCCATCTCTCCCATTATCGCTTCCACCAGAGCAACTTGCACAATTGGCCACTCTT 2496

Qy        730 ------AlaSerAsnAsnPro-----------------------AsnTyrThrSerGly 739
                  |||  |||                    |||     ||||||
Db       2497 CTTGCACAACAAAACCAACCAGGAAAAGAGCCTGTGGACAGCTTAAACAAAGAATCAGGA 2556

Qy        740 MetVal---------HisGlyAsn---------MetGlnTyrGlnSerGlnSerValAsn 753
                 :::         ||||||:::         |||   :::   |||   |||:::
Db       2557 TTCATACGGAATCCACATGGACATTCTTCAATGATGCCACACAGCTCAGGTTCTATC--- 2613

Qy        754 MetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGln 773
              |||       :::   ||||||      |||
Db       2614 ---CCTGTCCAAAACTCATTGCCACCTGCTCCG----------------------CCA 2646

Qy        774 GlySerSerAsnHisProValSerGlnProMetValGln--------------------- 786
              :::|||       |||     |||   ||||||
Db       2647 TCTGCATCACAGTTACAGGTTCATGCGCCACCAGTTCAAGGTTCAGTGCCACCAAATCCA 2706

Qy        787 ------GlnTyrGlnProGluAlaSerMetProAsnGlnAsnTyrGlyProIleProSer 804
                  :::   |||   |||   ||||||:::   |||      |||:::|||
Db       2707 TCCATTATGCATACACCGAATGCTCCTATGCCTTCTCACAACACTTTGCCTTTACCCCCT 2766

Qy        805 TyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSer 824
                 :::                  |||   |||   :::      |||
Db       2767 ATGCATCCTTCGGGA-----------------AATCCAGCTCATTCTTCCATGCCTTTG 2808

Qy        825 GlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGluProGlnAsnGln 844
              :::           |||                     :::      :::
Db       2809 AGATCGTTCGTCCCTCCTCTTCCTGAAGGCCCTCCACCCCTTAGGCAGCACACATCAAGT 2868

Qy        845 AlaLeuArgLeuGlnProMetIleSerGlyAspGlyGln------------------Gly 858
              ||||||:::   ||||||   :::            |||
Db       2869 GCTCTACAAGCGCAACCTGCACTTCCCTCTGGACCACAGACTAGCCAGCAACCGTCTGCT 2928

Qy        859 ThrThrAspGlyGluValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAla 878
                 |||      |||      :::|||  |||:::|||||||||   ||||||
Db       2929 CAGGAAGACCATCACGGAGACCCTCAAAAGCGCCTTCAAGCAACATTGCAATTGGCAGCA 2988

Qy        879 AsnLeuLeuLeuGlnIleGlnGlnLysGlnGlnGlnSerSerGlyThrProAlaGly 898
              ||||||  ||||||||||:::   :::                |||  |||
Db       2989 ACCCTACTTCAGCAGATACAGCAACAATCTAAA-----------------CCTGGTGGC 3030

Qy        899 Gln 899
              |||
Db       3031 CAG 3033

RESULT 7
US-10-424-599-59704
; Sequence 59704, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 59704
;   LENGTH: 1101
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_24924C.1
US-10-424-599-59704

Alignment Scores:
Pred. No.:              5.64e-51       Length:        1101
```

```
Score:                  669.00         Matches:       169
Percent Similarity:     54.57%         Conservative:  40
Best Local Similarity:  44.13%         Mismatches:    77
Query Match:            13.93%         Indels:        97
DB:                     17             Gaps:          15

US-09-920-705-3 (1-901) x US-10-424-599-59704 (1-1101)

Qy        282 GluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArgThr 301
              :::|||||||||                        ||||||:::
Db         54 CAGTTACCACCT-----------------------AGTGGTATT--------------- 74

Qy        302 AspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIleProGly 321
                 |||      |||  |||                ||||||||   |||:::
Db         75 ---ATGGGACCAAACATACCAATGCGACCTTTTGGTCCTCACAGT---GGTGTT------ 122

Qy        322 SerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAsp 341
                                          |||        :::|||   |||:::|||:::
Db        123 -------------------------GAATCTGTTATTTCTGGCCCTGAGTTTAATGAG 155

Qy        342 Val----------------------------------ValGlyLysGluProAsnTrp 349
              :::                                 |||||||||
Db        156 ATTAATGCGCTCCACAAATTTCAGGATGGTAGCTCCAAGAGTAGCATGGGTCCAAACTGG 215

Qy        350 ArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeu 369
              :::|||                 |||||||||   |||||:::|||
Db        216 AAAAGG-------------------------CCATCTCCTCCTGCACCAGGGATGCTT 248

Qy        370 ProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrpGluGly 389
                 |||||||||   |||   |||   |||   ||||||         :::|||:::
Db        249 TCTTCTCCTGCACCGGGTGCTAGGCTTCCTACAAGATCAACCTCGGGTGCATGGGATGTG 308

Qy        390 TyrAsp------------------ProAlaGlnLeuVal----ArgGluSerLysArgTh 402
                 |||                  :::   ||||||    |||:::||||||||::
Db        309 CTTGACATAAACCTATCATTCAGGAAGAGCACCTTGGTCTCCAAGAGATTCTAAACGTTC 368

Qy        402 rArgArgAspGlySerVal---------AspGlyPheThrProMet-------------- 414
              :|||  ||||||  :::          :::|||     |||:::
Db        369 AAGGATAGATGGACCCTTGCCTGTTGATGAAGGCCCATTTCCTTTGAGGAATATAGATGA 428

Qy        415 ---------------------------GlyValAspGluArgSerPheGlyArgGlySe 425
                                        |||:::|||          |||   |||||
Db        429 TCGGGGGTTAGCTCTAGAGCAGACATATGGGATTGACCCAGCTATTGATGGAGGTGGTTC 488

Qy        425 r----------------------------------------ValAlaAlaArgProIl 431
              |                                     |||:::::::|||
Db        489 TGGTCCATATGTAAACATTCAAGGAAAGAGTCACCTTGGTCCAGTTAGCTCAAGGATTAC 548

Qy        431 eArgGly--------------ProProAspSerAspHisIleTrpArgGlyMetIleAl 446
                 |||                |||||   |||||||||||||||||:::|||||
Db        549 AGCTGGAGTACATGGTGTCGCCCAACCTGATATTGATCATATTTGGCGTGGAGTTATTGC 608

Qy        446 aLysGlyGlyThrProValCysCysAlaArgCysValProMetGlyLysGlyIleGluTh 466
              ||||||||||||||||||||||   ||||||||||||||:::|||||||||||   ||
Db        609 AAAAGGGGGAACTCCTGTTTGCCGTGCTAGATGTGTACCAATTGGGAAAGGAATTGGGAC 668

Qy        466 rLysLeuProGluValValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHi 486
              |:::||||||:::||||||:::|||||||||||||||    |||::::::|||    |||||
Db        669 TGAGCTTCCTGATGTTGTTGATTGCTCAGCTAGGACGGGATTGGATATACTCACAAAACA 728

Qy        486 sTyrAlaValAlaIleGlyCysGluIleValPhePheValProAspArgGluGluAspPh 506
              |||||||   |||||||||  :::|||||||||||:::||||||   |||:::|||||
Db        729 TTATGCTGATGCAATTGGTTTTGATATTGTTTTCTTTCTGCCTGATAGTGAAGACGATTT 788

Qy        506 eAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLy 526
              |||||||||||||||||||||||||||||||:::|||:::|||||||||||||||||||
Db        789 TGCTTCATATACCGAGTTCCTTCGCTATCTTAGTGCGAAAAATCGTGCTGGTGTTGCCAA 848

Qy        526 sLeuAspAspGlyThrThrLeuPheLeuValProProSerAspPheLeuThrAspValLe 546
              |      |||   |||||||||||||||||||||||||||||||||||||||   |||||
Db        849 ATTTGTTGATAACACCACCTTATTCTTGGTGCCTCCTTCTGATTTTCTCACAAGAGTTTT 908

Qy        546 uGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLysLeuProProProAlaVa 566
```

```
                |:::||||||      ||||||||||||||||||||||||||||   ||||||      ||
Db        909 GAAAGTCACTGGACCTGAGCGTCTGTATGGTGTGGTTCTTAAGTTTCCACCA------GT 962

Qy        566 lProValThrAlaSerTyrArgGlnGluSerGlnSerAsn------ProLeuHisTyrMe 584
              ||||   :::|||      :::|||   |||                           |||||
Db        963 GCCAAGTAGTGCACCTATGCAACAACCATCACATTTGCGTGTACCAACTACTCAGTATAT 1022

Qy        584 tAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGluAs 604
              |   |||      :::   |||:::        :::   |||  |||  :::|||
Db       1023 GCAGCAG------AATCCTCCTTCGCAGACTGAATATGGTTTGATTCCTGTAAAAGAGGA 1076

Qy        604 nTyrIle 606
               :::|||
Db       1077 ACATATT 1083

RESULT 8
US-09-920-705-6
; Sequence 6, Application US/09920705
; Publication No. US20030079252A1
; GENERAL INFORMATION:
;  APPLICANT: Amasino, Richard M.
;  APPLICANT:  Schomburg, Fritz M.
;  APPLICANT:  Michaels, Scott D.
;  APPLICANT:  Patton, David
;  TITLE OF INVENTION: Floral Induction Gene
;  FILE REFERENCE: 960296.97214
;  CURRENT APPLICATION NUMBER: US/09/920,705
;  CURRENT FILING DATE:  2001-08-02
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 3715
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   OTHER INFORMATION: FPA promoter plus intron
US-09-920-705-6

Alignment Scores:
Pred. No.:              1.09e-48        Length:         3715
Score:                  651.50          Matches:        171
Percent Similarity:     27.23%          Conservative:   0
Best Local Similarity:  27.23%          Mismatches:     0
Query Match:            13.57%          Indels:         457
DB:                     10              Gaps:           1

US-09-920-705-3 (1-901) x US-09-920-705-6 (1-3715)

Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1832 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 1891

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1892 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 1951

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1952 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 2011

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2012 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 2071

Qy         81 SerGlnIleLysIleGluTyrAlaArgPro---------------------------- 90
              ||||||||||||||||||||||||||
Db       2072 AGTCAAATTAAGATCGAATACGCACGACCGGTTTGTTCTTATCTATATCTTCGTTTGTTC 2131

Qy         90 ------------------------------------------------------------ 90

Db       2132 TCTAACTTTGATTGTCTTTTGTCAACGATTATACTCTTTTTGCGAATTCATAGTCCAGGT 2191

Qy         90 ------------------------------------------------------------ 90

Db       2192 TCACAAAACTTTGATGATGCTTGTTTAGTCCAAAAATTTCTTGTTGAATCTGTTTTTTTT 2251

Qy         90 ------------------------------------------------------------ 90

Db       2252 TTCCTCATCGTACAAATCAAAGTCGAAACCTAGTTTTTTTCTATTATACGTCGTTAGCTT 2311

Qy         90 ------------------------------------------------------------ 90

Db       2312 AAGGCGAAACCTGATCCGATCGAAACGTCTTTTCTCAAATTACTTTGGTTATATCGAACT 2371

Qy         90 ------------------------------------------------------------ 90

Db       2372 CGCGCAAAGCCAAACCACAGAGAAGCTCTGCAAAATTTGATGTTAAAGCATATATAACTC 2431

Qy         90 ------------------------------------------------------------ 90

Db       2432 TTAGCGAATGAGCTCTGCAAAAGAATACATCAAACACATGTTTACTCTCGTTTATGCGAA 2491

Qy         90 ------------------------------------------------------------ 90

Db       2492 GAAGGTTTAATCCGATTGTCGTTTATCTGCAAACTTGTATCCGCGTTTAGCAGTCTGATT 2551

Qy         90 ------------------------------------------------------------ 90

Db       2552 TCAGGTTCGCTTCAGATGTTAAATCTCACAAGCTTGAGTATGAATGTATTGCGTACTCCA 2611

Qy         90 ------------------------------------------------------------ 90

Db       2612 CTCGAAATTCGCAATGGTGAAAATTTGGAAGCGAGCAAACTCTATCATCGGCCAAACAGA 2671

Qy         90 ------------------------------------------------------------ 90

Db       2672 TAAGAAATTTTGGAGTTTAAAAGTTCAGTTTCCTGCAAAAATCAAACCGCGTGGAGAAAT 2731

Qy         90 ------------------------------------------------------------ 90

Db       2732 TTGTCTATGGCGGTACGTAGATATCAATAACTGCATCTGCGAAACAGAAAATTCTGGCAA 2791

Qy         90 ------------------------------------------------------------ 90

Db       2792 CCGTATCGTCCTAACTCCTAGTATCGCTGGCACATATCCATATCAGTGTGAGGGAGTTTT 2851

Qy         90 ------------------------------------------------------------ 90

Db       2852 GGGCTTGTTTTGGCTCGATGCTTCAGGAAGACAAATTACGTGGTTGTTAAGGCGGCTAAC 2911

Qy         90 ------------------------------------------------------------ 90

Db       2912 TCTACCAATCAGAAACGCTTTATTCGAAGAACCATGTTTGTTCCTCAATTCCCATCCCTA 2971

Qy         90 ------------------------------------------------------------ 90

Db       2972 CGTACAATCTGGGCTTTCCCATTGTAGTTCCTTAGGAAGTTGACTTCTTCACAACCATTC 3031

Qy         90 ------------------------------------------------------------ 90

Db       3032 TGGGATTGGTACAGTTGCAGTGGACAAGACAAATATTCATTTGCAGGCAGACTCAAATTT 3091

Qy         90 ------------------------------------------------------------ 90

Db       3092 CAATGTCTTGCCGGCTTTGAGTACTTAAATGGAGTGTTCAGGGATTGGTTTATTTGGGAC 3151

Qy         90 ------------------------------------------------------------ 90

Db       3152 TCAGGCAATGGATAAAGACAGGAATGTTTTGCAGAAAGTATTAATGTCTTTCCGGCTTTG 3211

Qy         90 ------------------------------------------------------------ 90

Db       3212 GGGACTACATACTTCTCCTACAGACAAAATCTGATGTTTAGGCAAGAGAAACTATTAACA 3271

Qy         90 ------------------------------------------------------------ 90
```

```
Db       3272 CTGATTTAATAGAGAAAGAGGAGATGGTTTCTCCTTGCGGCAATTTTATTTGTTTAGGAA 3331

Qy         90 ------------------------------------------------------------ 90

Db       3332 AGCAATTGATATGAATTGGTGTCGTAGTGTAGTTGAAATTACTAGTTAGTTTGTGTGTTT 3391

Qy         90 ------------------------------------------------------------ 90

Db       3392 AGTTTCCTTGATGTTTGATGCTTTATTCTTGGCAACCTATCTGGGTAGTATCGCCTTCTT 3451

Qy         91 ---------------------AlaLysProCysLysSerLeuTrpValGlyGlyIleGly 103
                                   |||||||||||||||||||||||||||||||||||||||
Db       3452 ATCGACCTTTTCTTGTTGCAGGCAAAACCTTGTAAGAGTCTATGGGTGGGTGGAATCGGC 3511

Qy        104 ProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAsp 123
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3512 CCTAATGTCTCCAAGGATGACCTGGAGGAAGAGTTCAGCAAGTTTGGGAAAATCGAGGAT 3571

Qy        124 PheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAla 143
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3572 TTTAGGTTTCTCAGAGAACGCAAGACAGCTTTCATTGATTATTATGAGATGGATGATGCT 3631

Qy        144 LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPhe 163
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3632 TTACAGGCTAAGAGCATGAATGGAAAGCCTATGGGTGGTAGCTTTTTGCGTGTTGATTTT 3691

Qy        164 LeuArgSerGlnAlaProLysLys 171
               |||||||||||||||||||||||
Db       3692 CTCCGGTCACAAGCGCCAAAAAAA 3715

RESULT 9
US-10-424-599-65572
; Sequence 65572, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 65572
;   LENGTH: 801
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(801)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_30226C.1
US-10-424-599-65572

Alignment Scores:
Pred. No.:              1.09e-35        Length:         801
Score:                  499.00          Matches:        105
Percent Similarity:     65.55%          Conservative:   32
Best Local Similarity:  50.24%          Mismatches:     63
Query Match:            10.39%          Indels:         9
DB:                     17              Gaps:           3

US-09-920-705-3 (1-901) x US-10-424-599-65572 (1-801)

Qy          6 LysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeu 25
              |||          ::::::   |||      :::|||||||||||||||||:::|||
Db        165 AAATCCGTGGAAGGCGAGGAGTGGGGAACCCCCACCAACAACCTCTGGGTCGGGAATCTG 224

Qy         26 ThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAsp 45
                  |||   |||:::|||||||||   |||||||||       ||||||   :::|||

Db        225 CCGACGGAGGTCACAGATTCCGATCTCATGGAACTATTCGCTCCGTACGGTTCTCTGGAC 284

Qy         46 ArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyrArgHisValGluGlu 65
                 :::      ||||||||||||   |||||||||:::   :::|||   :::|||:::
Db        285 TCACTCATCTCCTACTCCTCGCGCACCTTCGCTTTCGTCTTGTTCAGGCGCATCGAAGAC 344

Qy         66 AlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleLysIle 85
              |||   |||||||||       ||||||||||||   |||   |||   ||||||:::|||
Db        345 GCTAAGGCAGCGAAATCCAACTTGCAAGGCGCGTTGCTGCGCGGGTTTCAAATCAGAATC 404

Qy         86 GluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGlyGlyIleGlyProAsn 105
              |||:::||||||||||||||||||||||||   ||||||||||||||      |||
Db        405 GAGTTTGCCAGACCGGCAAAGCCGTGCAAACAATTGTGGGTGGGTGGTTTTAGTCCCACT 464

Qy        106 ValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArg 125
              |||:::::::::|||||||||   ||||||   |||||||||||||||||||||||:::
Db        465 GTTGCAAGGGAAGATTTGGAAGCGGAATTTCGCAAATTCGGTAAGATTGAGGATTTTAAG 524

Qy        126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
              |||      :::|||  ||||||   :::::::::      :::|||  |||   :::
Db        525 TTCTTTATAGACCGCGGCACTGCGTGTGTTGAGTTTCTCAATCTGGATGCTGCTGCTCGG 584

Qy        146 Ala---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeu 164
              |||   |||  ||||||||||||      ||||||      :::   ||||||||||||
Db        585 GCGATGAAAGTCATGAACGGGAAGCGGNTAGGTGGACGCCAGATTTGTGTAGATTTCCTT 644

Qy        165 ArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsnGlyAsn 184
              |||||||||:::   :::::::::               :::         :::|||
Db        645 CGATCACAATCTACGAGAAGAGAT--------------TTTTTGGTTGATCATGGGCAG 689

Qy        185 MetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspValGlnPro 204
                     :::||||||   |||              :::   ||||||
Db        690 TTTCAGGCTAGACCACAG------CATTTACAACCTTCAATTGGGAGGAATAATCAACCC 743

Qy        205 SerLysValLeuTrp-IleGlyPhe 212
              ||||||   |||||| |||||||||
Db        744 AGTAAAAATCTGTGGGATAGGGTTT 768

RESULT 10
US-09-920-705-5
; Sequence 5, Application US/09920705
; Publication No. US20030079252A1
; GENERAL INFORMATION:
;  APPLICANT: Amasino, Richard M.
;  APPLICANT:  Schomburg, Fritz M.
;  APPLICANT:  Michaels, Scott D.
;  APPLICANT:  Patton, David
;  TITLE OF INVENTION: Floral Induction Gene
;  FILE REFERENCE: 960296.97214
;  CURRENT APPLICATION NUMBER: US/09/920,705
;  CURRENT FILING DATE:  2001-08-02
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 349
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   OTHER INFORMATION: Portion of FPA coding region
US-09-920-705-5

Alignment Scores:
Pred. No.:              3.68e-33        Length:         349
Score:                  466.00          Matches:        92
Percent Similarity:     94.95%          Conservative:   2
Best Local Similarity:  92.93%          Mismatches:     3
Query Match:            9.71%           Indels:         2
DB:                     10              Gaps:           1

US-09-920-705-3 (1-901) x US-09-920-705-5 (1-349)

Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 60

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40

Db         61 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 120

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60

Db        121 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 180

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80

Db        181 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 240

Qy         81 SerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLysSerLeuTrpVal 99

Db        241 AGTCAAATTAAGATCGAATACGCACGACCGGTT------TGTTCTTATCTATATCTT 291

RESULT 11
US-10-425-114-29888
; Sequence 29888, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 29888
;   LENGTH: 1630
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: UC-GMROPIC022E10_FLI
US-10-425-114-29888

Alignment Scores:
Pred. No.:                5.24e-32        Length:          1630
Score:                    463.00          Matches:         155
Percent Similarity:       43.09%          Conservative:    60
Best Local Similarity:    31.06%          Mismatches:      194
Query Match:              9.64%           Indels:          90
DB:                       17              Gaps:            17

US-09-920-705-3 (1-901) x US-10-425-114-29888 (1-1630)

Qy        432 ArgGlyProProAspSerAspHisIle-----------------TrpArgGlyMetIle 445

Db         49 AGATTCACGCCTGAATCCGATAGTTCATCTTTAACTGAGTGGAAATGGGAAGGAACCATT 108

Qy        446 AlaLysGlyGlyThrProValCysCysAlaArgCysValProMetGlyLysGlyIleGlu 465

Db        109 GCTAAAGGCGGAACCCCTGTTTGTTGTGCACGCTGCTTCCCTGTTGGGAAAGTCCTAGAT 168

Qy        466 ThrLysLeuProGluValValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLys 485

Db        169 ATGATGTTACCTGAGTTCTTGGATTGCACTGCAAAAACTGGTCTAGATATGCTTTCAAAG 228

Qy        486 HisTyrAlaValAlaIleGlyCysGluIleValPhePheValProAspArgGluGluAsp 505

Db        229 CATTACTACCAAGCAGTTGGTGTTTGGGTTGTTTTCTTTGTGCCTGGAAGTGATGCTGAC 288

Qy        506 PheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAla 525

Db        289 ATGGAATGCTACAATGAATTTATGCATTATCTGGAGGAAAAGCAGCGCGCTGCAGTTGCC 348

Qy        526 LysLeuAspAspGlyThrThrLeuPheLeuValProProSerAspPheLeuThrAspVal 545

Db        349 AAGTTGGATGACAAGACCACCTTATTTCTTGTTCCTCCATCAGAGTTCTCAGAGAAAGTA 408

Qy        546 LeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLysLeuProProProAla 565

Db        409 CTGAAGGTACCTGGTAGGTTGAGCATATCTGGTGTTATTCTGAGGTTGGAGAATCCTGAT 468

Qy        566 ValProValThrAlaSerTyr------ArgGlnGluSerGlnSerAsnProLeuHisTyr 583

Db        469 TTAAATCATGGTCCTGAACACATTCAAAGAGAAATGACTAACAAAAATCTGCTGTCCTAT 528

Qy        584 MetAspGlnAlaArgAspSerProAlaAsnAlaSerHis------SerLeuTyrProPro 601

Db        529 AATGAG--------------------AATATACTGCATCTGAAGTCTACATTTCCTTCA 567

Qy        602 ---ArgGluAsnTyrIleArgGlyAlaProGluHisLeuThrAla--------------- 615

Db        568 GTACGTGTACCTACATTCCCATCTATTCCAGAAATGGGTAATTCTGGAATCAGCAATCTT 627

Qy        616 --------------------AlaSerLysProSerValSerGluProLeuArgIlePro 628

Db        628 TCATTTCTGGGAAATAAATTTGCAGCAGCTCCATCAGTTTCTGATTCAGCTCGT------ 681

Qy        629 AsnAsnAlaAlaProGlnAlaGlyValSerLeuThr----------------------- 640

Db        682 -----------------GCTGTGGCTAGCATGTCCGAGTTTCATGATGAAAGAAGTCAC 723

Qy        641 ------ProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAla 658

Db        724 AACTATCCTACCCAGCAGCGTACTTCTTCTCAAAATCTCCAGAACTTCTCTAATAAAGCT 783

Qy        659 AlaProGluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGln 678

Db        784 CTACCA------TTGCAGCCATCAAGTGGTGCTGGAGAACCTATTGCTGATGAACACCAA 837

Qy        679 SerAsnGlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThr 698

Db        838 --------------------------------CCTATCATTCCAAGGGCAGCGCAGGAT 864

Qy        699 Val------------HisAspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyr 714

Db        865 GTAAACGCGATTCAACATCCCAGTGGAATTTCTGGTATTCCATTTTATGGTGAT------ 918

Qy        715 ThrProAlaGlyGlnLeuProProProProSerArgTyr-----------------Pro 728

Db        919 ---------AGCAAGTTGTCTTATCCAGATATAAGACATTTAGATCCTTTATCAGTGCCT 969

Qy        729 ProAlaSerAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMetGlnTyrGln 748

Db        970 GTGGGAGCCCTTGCTCCAGAACAGCTTGCACAACTAGCAGCAACGCTTCTTGAGCAGCAG 1029

Qy        749 SerGlnSerValAsnMetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyr 768

Db       1030 AGGCAATCGGGAGGCTCTTCGAGTACATCTGCGTTGGCTGAT---CCCCAGCAGATTAAC 1086

Qy        769 SerMetTyrThrGlnGlySerSerAsnHisProValSerGlnProMetValGlnGlnTyr 788

Db       1087 AGGTTTGGCACATCCGACACCTCATCCAGATTTATCTCA---------------------- 1125

Qy        789 GlnProGluAlaSerMetProAsnGlnAsnTyrGlyProIleProSerTyrGlnGlnAla 808

Db       1126 TCCGACAACTCATCCAGACCACCCCAGAAATATGCTACAGAAAACAATGTGGTGAATTCT 1185

Qy        809 AsnPheHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAla 828

Db       1186 GATCTCTCAGCATCTCAGATGGGTCAAAGTCAAATGCTACAGATGCAAAAGCAACAACAG 1245

Qy        829 AlaMetGlnProPro------AlaAspLysAlaAsnLeuGluProGlnAsnGlnAlaLeu 846

Db       1246 ATCGTAAATGTGCCTCAGTTGTCACAGATAGTTCAAAGGGAACCACAAAGTGAGGCTAAT 1305

Qy        847 ArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThrAspGlyGluValAspLys 866
```

```
Db       1306 GGAAACCAGTTG-----------GATAGCAGCTTGCAGGAAGATGCTGATGTAGATCCA 1353

Qy        867 AsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsnLeuLeuLeuGlnIleGln 885
Db       1354 CAGAGACGTCTACAAGCAACACTGCAGCTGGCGGCTGTTCTTCTTCACCAAATCCAG 1410
```

RESULT 12
US-10-424-599-40095
; Sequence 40095, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 40095
; LENGTH: 1631
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_136205C.1
US-10-424-599-40095

Alignment Scores:

Pred. No.:	1.5e-30	Length:	1631
Score:	447.00	Matches:	155
Percent Similarity:	42.80%	Conservative:	59
Best Local Similarity:	31.00%	Mismatches:	195
Query Match:	9.31%	Indels:	91
DB:	17	Gaps:	17

US-09-920-705-3 (1-901) x US-10-424-599-40095 (1-1631)

```
Qy        432 ArgGlyProProAspSerAspHisIle-----------------TrpArgGlyMetIle 445
Db         49 AGATTCACGCCTGAATCCGATAGTTCATCTTTAACTGAGTGGAAATGGGAAGGAACCATT 108

Qy        446 AlaLysGlyGlyThrProValCysCysAlaArgCysValProMetGlyLysGlyIleGlu 465
Db        109 GCTAAAGGCGGAACCCCTGTTTGTTGTGCACGCTGCTTCCCTGTTGGGAAAGTCCTAGAT 168

Qy        466 ThrLysLeuProGluValValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLys 485
Db        169 ATGATGTTACCTGAGTTCTTGGATTGCACTGCAAAAACTGGTCTAGATATGCTTTCAAAG 228

Qy        486 HisTyrAlaValAlaIleGlyCysGluIleValPhePheValProAspArgGluGluAsp 505
Db        229 CATTACTACCAAGCAGTTGGTGTTTGGGTTGTTTTCTTTGTGCCTGGAAGTGATGCTGAC 288

Qy        506 PheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAla 525
Db        289 ATGGAATGCTACAATGAATTTATGCATTATCTGGAGGAAAAGAAGCGCGCTGCAGTTGCC 348

Qy        526 LysLeuAspAspGlyThrThrLeuPheLeuValProProSerAspPheLeuThrAspVal 545
Db        349 AAGTTGGATGACAAGACCACCTTATTTCTTGTTCCTCCATCAGAGTTCTCAGAGAAAGTA 408

Qy        546 LeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLysLeuProProProAla 565
Db        409 CTGAAGGTACCTGGTAGGTTGAGCATATCTGGTGTTATTCTGAGGTTGGAGAACCCTGAT 468

Qy        566 ValProValThrAlaSerTyr------ArgGlnGluSerGlnSerAsnProLeuHisTyr 583
Db        469 TTAAATCATGGTCCTGAACACATTCAAAGAGAAATGACTAACAAAAATCTGCTGTCCTAT 528

Qy        584 MetAspGlnAlaArgAspSerProAlaAsnAlaSerHis------SerLeuTyrProPro 601
```

```
Db        529 AATGAG--------------------AATATACTGCATCTGAAGTCTACATTTCCTTCA 567

Qy        602 ---ArgGluAsnTyrIleArgGlyAlaProGluHisLeuThrAla--------------- 615
Db        568 GTACGTGTACCTACATTCCCATCTATTCCAGAAATGGGTAATTCTGGAATCAGCAATCTT 627

Qy        616 ---------------------AlaSerLysProSerValSerGluProLeuArgIlePro 628
Db        628 TCATTTCTGGGAAATAAATTTGCAGCAGCTCCATCAGTTTCTGATTCAGCTCGT------ 681

Qy        629 AsnAsnAlaAlaProGlnAlaGlyValSerLeuThr----------------------- 640
Db        682 ------------------GCTGTGGCTAGCATGTCCGAGTTTCATGATGAAAGAAGTCAC 723

Qy        641 ------ProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAla 658
Db        724 AACTATCCTACCCAGCAGCGTACTTCTTCTCAAAATCTCCAGAACTTCTCTAATAAAGCT 783

Qy        659 AlaProGluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGln 678
Db        784 CTACCA------TTGCAGCCATCAAGTGGTGCTGGAGAACCTATTGCTGATGAACACCAA 837

Qy        679 SerAsnGlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThr 698
Db        838 --------------------------------CCTATCATTCCAAGGGCAGCGCAGGAT 864

Qy        699 Val------------HisAspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyr 714
Db        865 GTAAACGCGATTCAACATCCCAGTGGAATTTCTGGTATTCCATTTTATGGTGAT------ 918

Qy        715 ThrProAlaGlyGlnLeuProProProProSerArgTyr------------------Pro 728
Db        919 ---------AGCAAGTTGTCTTATCCAGATATAAGACATTTAGATCCTTTATCAGTGCCT 969

Qy        729 ProAlaSerAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMetGlnTyrGln 748
Db        970 GTGGGAGCCCTTGCTCCAGAACAGCTTGCACAACTAGCAGCAACGCTTCTTGAGCAGCAG 1029

Qy        749 SerGlnSerValAsnMetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyr 768
Db       1030 AGGCAATCGGGAGGCTCTTCGAGTACATCTGCGTTGGCTGAT---CCCCAGCAGATTAAC 1086

Qy        769 SerMetTyrThrGlnGlySerSerAsnHisProValSerGlnProMetValGlnGlnTyr 788
Db       1087 AGGTTTGGCACATCCGACACCTCATCCAGATTTATCTCA---------------------- 1125

Qy        789 GlnProGluAlaSerMetProAsnGlnAsnTyrGlyProIleProSerTyrGlnGlnAla 808
Db       1126 TCCGACAACTCATCCAGACCACCCCAGAAATATGCTACAGAAAACAATGTGGTGAATTCT 1185

Qy        809 AsnPheHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAla 828
Db       1186 GATCTCTCAGCATCTCAGATGGGTCAAAGTCAAATGCTACAGATGCAAAAGCAACAACAG 1245

Qy        829 AlaMetGlnProPro------AlaAspLysAlaAsnLeuGluProGlnAsnGlnAlaLeu 846
Db       1246 ATCGTAAATGTGCCTCAGTTGTCACAGATAGTTCAAAGGGAACCACAAAGAGAGGCCAAT 1305

Qy        847 ArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThrAspGly-GluValAspLy 866
Db       1306 GGAAACCAGTTG-----------GATAGCAGCTTGCAGGAAGATGACTGATGTAGATCC 1353

Qy        866 sAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsnLeuLeuLeuGlnIleGln 885
Db       1354 ACAGAGACGTCTACAGGCAACACTGCAGCTGGCGGCTGTTCTTCTTCACCAAATCCAG 1411
```

RESULT 13
US-10-425-114-31541
; Sequence 31541, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong

```
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 31541
;   LENGTH: 2159
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: UC-ZMFLB73186F01_FLI
US-10-425-114-31541

Alignment Scores:
Pred. No.:                1.16e-26      Length:          2159
Score:                    406.00        Matches:         174
Percent Similarity:       35.84%        Conservative:    79
Best Local Similarity:    24.65%        Mismatches:      204
Query Match:              8.46%         Indels:          250
DB:                       17            Gaps:            29

US-09-920-705-3 (1-901) x US-10-425-114-31541 (1-2159)

Qy        360 LeuProSerProThrGly----------ProGlyIleLeuProSerProAlaGlnGly 375
              ||||||   |||                |||          :::      :::|||
Db         44 TTGCCGCCCGCCCGCCGCTCGCTGCTGCCCCGCCGCTCGCCGCGGCCAATGATGGCGAGGGG 103

Qy        376 ThrArgArgPro--------------MetArgSerAsnProAspSerTrpGluGlyTyr 390
                  ||||||                :::|||      |||
Db        104 CCGCCGCGGCCCGGGTGGACGGGGGAGGTTCGGCGGCGGCCCGCTGGGCGACGAGCGGCC 163

Qy        391 AspProAlaGlnLeuValArgGluSerLysArgThrArgArgAspGlySerValAspGly 410
                 |||                 |||:::|||   ||||||         |||
Db        164 AGTCCG-----------------TCACGGCGTCCGCGTCGAGGAGAACCCGTACTTCCC 205

Qy        411 PheThrProMetGlyValAspGluArgSerPheGlyArgGlySerValAlaAlaArgPro 430
                 |||      ||||||                |||  |||:::     |||  |||
Db        206 ATTACCGCGTCGGGCGTC---------GGGCTGGGGAGTGGCACCGCCGTCGCGGCACCT 256

Qy        431 IleArgGlyProProAspSerAspHisIleTrpArgGlyMetIleAlaLysGlyGlyThr 450
                 :::   |||                     |||  |||              |||
Db        257 GTGGGTGGG--------------------TGGCCTGGCGCCCAGCGTCACGGCGTCCGA 295

Qy        451 ProVal---------------------------------CysCysAlaArgCysValPro 459
                 ||||||                               ||||||   |||      |||
Db        296 CCTGTCGGAGCTGTTCCTCCGGTGCGGCGACGTCGAGGGTGTTGCCCGCGAACCCGGCCG 355

Qy        460 ------------------------------------MetGlyLysGly--------- 463
                                                  |||:::|||
Db        356 GAACTTCGCGTTCGTGAGCTTCCTGCGGGAGGGGGACGCTGTGGCCGCGGTGCGGGAGCT 415

Qy        464 ---------------------------------IleGluThrLysLeuProGlu 470
                                                  :::      :::   ||||||
Db        416 CCAGGGGGCTCGCCTGGCTGGGGCGCCTGTTAGGATCGAGTTTT-CCAAGGGGGCCTGAG 474

Qy        471 ValValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAla 490
                 ::::::|||:::|||||||||   |||:::||||||:::|||||||||      |||
Db        475 TTTCTGGATTGCACTGCTAGGACAAGTCTAGATATGCTTTCTAAGCACTACTATCAAGCT 534

Qy        491 IleGlyCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrThr 510
                 |||      :::|||||||||||||||:::   :::   |||      :::|||
Db        535 ACTGGCAGCTGGGTCGTATTTTTTGTCCCAGAAAATGATGCGGACATGACAGCTTATAAT 594

Qy        511 GluPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGly 530
                 :::|||:::   ||||||      |||   ||||||   |||   ||||||   :::
Db        595 GATTTCATGAGTTATCTAGGTGATAAGCAGCGTGCAGCAGTCTGTAAGCTTGGAGAGAGC 654

Qy        531 ThrThrLeuPheLeuValProProSerAspPheLeuThrAspValLeuGlnValThrArg 550
                 :::|||||||||||||||||||||||||||||||||          ||||||:::|||
Db        655 AGCACCTTGTTTCTTGTTCCACCTTCAGATTTCTCAGAGCAAGTACTGAGAGTGCCTGGA 714

Qy        551 GlnGluArgLeuTyrGlyValValLeuLysLeuProProProAlaValProValThrAla 570
                       :::   ||||||:::|||:::
Db        715 AATGTCAGCATATCTGGAGTCATTCTGAGGTTT-------------------------- 747

Qy        571 SerTyrArgGlnGluSerGlnSerAsnProLeuHisTyr------------------Met 584
                                ||||||||||||   :::                    :::
Db        748 ---------------CAGCAGTCAAATCCAGATTACAGCTCACCAAATCGCAAACCATTG 792

Qy        585 AspGlnAlaArgAspSerProAlaAsnAlaSerHis------------------------ 596
                 |||:::|||          |||::::::||||||:::
Db        793 GATAGAGCTCAC------CCATCTTCTGCAAGTAATTTTAACACTGATGTAAGAAATTGT 846

Qy        597 ----------------------------------------SerLeuTyrProProArg 602
                                                       |||   :::|||
Db        847 GAAGATCTGCATGCACTGAGAAGGCTCAATCCTTCAGATATCAGCACATTCCCTCAGGGT 906

Qy        603 GluAsnTyrIleArgGlyAlaProGluHisLeuThrAlaAlaSer--------------- 617
                  :::|||:::      :::          |||   |||:::
Db        907 CCAGATTATGTCTGCTCATCAGGTGGAAGCTATACCCCAGCAAATGCAGATTTTATTCCG 966

Qy        618 ------LysProSerValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAla 635
                    ||||||      :::   |||      :::|||  :::      :::
Db        967 CCCTATAAGCCAGAGAGTGCTCCTCCATATGTCGTCCCTCAATCACCGCAAGAAAGGCCG 1026

Qy        636 GlyValSerLeuThrProGluLeuLeuAla--------------ThrLeuAlaSerIle 650
                    |||   |||      :::|||                    |||   :::
Db       1027 CCAGCGGAGCTC---CCTATGGGTATTGCACAAGGCCAACACCAGCAACTCCCAAACAGG 1083

Qy        651 LeuPro---------------------------------------------------- 652
                 ||||||
Db       1084 CTGCCCTCAGGATGGTCCAATAACCTGGATAATCCAAATCCAAATCTAGGTTCTGGTAAC 1143

Qy        653 ---AlaThrSerGlnProAla---AlaProGluSerHisGlnProMetSerGlyProSer 670
                  :::   :::|||   |||  :::      :::      |||   :::
Db       1144 TTCAGTTTGGCACAGGGTGCAATCTCTAACAGGACACAGGATCCGTACACATTTGCTCCT 1203

Qy        671 ThrValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyGluAlaPro---Ser 689
                 ||||||||||||         ::::::|||      :::||||||||||      |||
Db       1204 GGAGTAGTATCAACTGGG---ACAGCTTCAGGGTATGCATCAGGTGAAGCATCCAACAGC 1260

Qy        690 GlnAlaTrpLysArgGlyProGlnThrValHisAspAlaSerAsnGlnSerPheGlnGln 709
                 :::|||                                    :::      |||
Db       1261 ATGTCCTGG-----------------------------------------GCTTCCATGCAA 1281

Qy        710 TyrGlyAsnGlnTyrThrProAlaGlyGlnLeuProProProProSerArgTyrProPro 729
                 |||  :::|||                    |||      |||          ||||||
Db       1282 TACAACTCACAGCAGGTAGCTAGATCTGATCAACCTACTATTCCAGTATCACTGCCACCT 1341

Qy        729 ---------------------------------------------------------- 729

Db       1342 GATCAACTTGCACAGCTAGCTGCCCTTCTTGCACGACAAAACCAACCTAGAAAAGTTGGT 1401

Qy        730 -------------------------------------AlaSerAsnAsnProAsnTyr 736
                                                     |||   ||||||:::
Db       1402 TTGCCTGTGGACAGCTCAAACAATCTGTCTGGATTCATACAGAGTCCCAATCCTCATGGA 1461

Qy        737 ThrSerGlyMetValHisGlyAsnMet------------------------------- 745
                 :::   |||:::   ||||||
Db       1462 CATGCTACAATGATGCCAGGCAACTCTGGCTCCATCCCTATCCAGAACTCACTGCCACCT 1521

Qy        746 ------GlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeuProAsnMet 763
                       |||            |||:::|||   :::||||||:::   ::::::
Db       1522 GTTCCATCTATGCAACAATTACATGCCAATGTGCCA---GTGTCACCAATAGGCTCGCTA 1578

Qy        764 ProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProValSerGlnPro 783
                 |||   |||   :::::::::   |||                    :::   |||
Db       1579 CCTGCAAATCCATTCAATGTGCCGACCAGTGCTCCTGTACTTTGTAACAGCACTTCGCCT 1638

Qy        784 MetValGlnGlnTyr-----------GlnProGluAlaSerMetProAsnGlnAsnTyr 799
```

```
Db       1639 ATGCTGCCCAGTCATGCTTTGGTCAACCCAGCTCATTCTTCCATGCCAATGAGACCGTTC 1698

Qy        800 ------------------GlyProIleProSerTyrGlnGlnAlaAsnPheHisGlyVal 813

Db       1699 ATTCCTCCACTTCCTGAAGGTCCT---CCACCCTTTCAGCAGCAAACATCAAGTGCTGTC 1755

Qy        814 ThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGlnProPro 833

Db       1756 ACA---------GTGCAGCCTTTAGGTCCTTCTAGCCAGCAGTCTAGCCAACAACTACTG 1806

Qy        834 AlaAspLysAlaAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSer 853

Db       1807 GCTCAGGACCTTGATGGAGATCCTCAA------------------------------- 1833

Qy        854 GlyAspGlyGlnGlyThrThrAspGlyGluValAspLysAsnGlnArgTyrGlnSerThr 873

Db       1834 ----------------------------------------AAGCGCCTTCAAGCAACA 1851

Qy        874 LeuGlnPheAlaAlaAsnLeuLeuGlnIleGlnGlnLysGlnGlnGlnGlnSerSer 893

Db       1852 CTGCAACTGGCAGCAACCCTGCTTAAGCAGATA-----------------CAAAATCAA 1893

Qy        894 GlyThrProAlaGlyGln 899

Db       1894 TCAAATCCTGGTGGCCAG 1911

RESULT 14
US-10-767-701-2544
; Sequence 2544, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 2544
;   LENGTH: 815
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS43199_1
US-10-767-701-2544

Alignment Scores:
Pred. No.:              4.94e-25        Length:         815
Score:                  382.00          Matches:        87
Percent Similarity:     61.21%          Conservative:   44
Best Local Similarity:  40.65%          Mismatches:     57
Query Match:            7.96%           Indels:         26
DB:                     18              Gaps:           8

US-09-920-705-3 (1-901) x US-10-767-701-2544 (1-815)

Qy         14 GlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThrGluSerAsp 33

Db        204 GGCCCGGAGACGAACACGCTCCGCGTGGGCAACCTGCCCTTGCACGTGACCGAGGGCGAT 263

Qy         34 LeuThrGluLeuPheGlyArgTyrGlyAspIleAsp---ArgIleThrValTyrSerSer 52

Db        264 TTGTTGGCGCTGTTCGGGCCGCACGGCGCGCTCGATTGCGCCCTCGCACGAGCCGGATCT 323

Qy         53 ArgGlyPheAlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAla 72

Db        324 CGCAGCTACGCGTTCCTCCTCTTCCGTTCCCCAGCCGAGGCCCGCGCTGCCGTCGAGGCC 383

Qy         73 LeuGlnGlyAlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAlaArgProAlaLys 92

Db        384 ACCCGGGGTGAAAAGGTCAAGGGGGCCGCCATGCGCACCGAGTTCGCGCGGCCGGCTAAA 443

Qy         93 ProCysLysSerLeuTrpValGlyGlyIleGlyProAsnValSerLysAspAspLeuGlu 112

Db        444 GCTGTTAGGAATCTGTGGGTTGGTAGCATTAGCCCATCGGTTTCAAAGGAGGAGCTTGAG 503

Qy        113 GluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLysThr 132

Db        504 GAGGAGTTTCAAAAGTTTGGAAAAGTTGAAGGTGTTGCATTCTCCCAAGATCAGACTTCG 563

Qy        133 AlaPheIleAspTyrTyrGluMetAspAspAlaLeuGlnAla---LysSerMetAsnGly 151

Db        564 GCGTACATTGATTTTGAGAAGCTTGAAGATGCAATTTCTGCTCACAGATCCTTGAATGGA 623

Qy        152 LysProMetGlyGlySerPheLeuArgValAspPheLeuArgSerGlnAlaProLysLys 171

Db        624 AAAACTTTAGGTGGCAAGGAATTGTGTGTTGATTTCCAGAGGTCCAAGGGA------AGA 677

Qy        172 GluGlnTrp------AlaGlySerTyrAspAsnArg---AsnGlyAsnMetAsnHisLys 188

Db        678 GCGGAATGGTCAGACGCTAGCAGCTTCAATGGTAGAGTATCAGGGCCAGTGGGTGATAAG 737

Qy        189 ------------ProGlnTyrProHisSerTyrGluAspPheLysGly------------ 200

Db        738 AGGGGCAGTGGCCCTCCA----------------------AAGGGTTCTGCTGGAATA 773

Qy        201 ---------AspValGlnProSerLysValLeuTrpIleGly 211

Db        774 CGGATGCGAGAAGCACAGCCTACCAATGTTCTCTGGGTTGGT 815

RESULT 15
US-10-767-701-30057
; Sequence 30057, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 30057
;   LENGTH: 606
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 9849454
US-10-767-701-30057

Alignment Scores:
Pred. No.:              4.14e-23        Length:         606
Score:                  359.00          Matches:        90
Percent Similarity:     52.61%          Conservative:   21
Best Local Similarity:  42.65%          Mismatches:     62
Query Match:            7.48%           Indels:         38
DB:                     18              Gaps:           5

US-09-920-705-3 (1-901) x US-10-767-701-30057 (1-606)

Qy        484 AlaLysHisTyrAlaValAlaIleGlyCysGluIleValPhePheValProAspArgGlu 503

Db          9 GCGAAGCACTATGCGGATGCCACTGGATTTGACATTGTCTTTTTCTTACCAGATAGTGAA 68

Qy        504 GluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAspArgAlaGly 523

Db         69 GATGACTTTGTTTCTTATACTGAGTTCTTGCGCTACCTGGGCTCAAAAAGCCGGGCAGGG 128

Qy        524 ValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProProSerAspPheLeuThr 543

Db        129 GTTGTAAAAGTTGATGCAGCTACTACTTTATTTTTGGTGCCACCATCAGATTTCTTAACA 188
```

```
Qy        544 AspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLysLeuProPro 563
              :::|||||||||||||          ||||||||||||||||||||||||   :::|||
Db        189 AATGTGTTACAAGTTGATGGTCCAGAGCGCCTTTATGGTGTGGTATTGCACATTCCACAA 248

Qy        564 ProAlaValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnProLeuHisTyr 583
                :::           |||           |||   :::   :::              |||
Db        249 ATCTCTGCTGCTGCTGCTGCACTGAGGCCACAGTTAACTGGGACAGAACAGCAACCTTAC 308

Qy        584 MetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGlu 603
                |||:::           |||           :::|||:::      |||
Db        309 TATGATGAAAGGGAAACTCTGCCTACACAAAGAAAGTATAGTATCATTTCTCCT------ 362

Qy        604 AsnTyrIleArgGlyAlaProGluHisLeuThrAlaAlaSerLysProSerValSerGlu 623
                                    ||||||   |||      :::   |||:::   |||
Db        363 ------------AATGGCAGTGGCCACCTTGATGCTGATTATCGGGCATCTTTGCATGAG 410

Qy        624 ---------------------------ProLeuArg----------------------- 626
                                          |||   |||
Db        411 GATTCAATGCAACGTTTGGGGCATATTCCTGGAAGGCCTCGGGTGGATGAAGGCCAAGCA 470

Qy        627 ---------------------IleProAsnAsnAlaAlaPro------GlnAlaGlyVal 637
                                   |||   |||   |||          |||
Db        471 GTCCAGCCAGCTCTTGCAGGTTTTCCTGCTAACCAAGCAACTGGGCTGCAAGTGCAATCT 530

Qy        638 SerLeuThrProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnPro 657
              ||||||   |||:::::::::|||||||||||   :::||||||:::
Db        531 TCACTCAAGCCTGACATTATGGCTACTTTAGCAAAACTTCTGCCTAGTGTG--------- 581

Qy        658 AlaAlaProGluSerHisGlnProMetSerGly 668
                  :::|||   |||   :::||||||
Db        582 ---------CAATCATCACAACTGGTTAGTGGT 605
```

Search completed: April 8, 2005, 16:35:33
Job time : 1297 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2005, 08:31:28 ; Search time 6977 Seconds
(without alignments)
4915.566 Million cell updates/sec

Title: US-09-920-705-3
Perfect score: 4801
Sequence: 1 MALSMKPFRADDSGFQSNNL.....LQIQQKQQQQSSGTPAGQGP 901

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09920705/runat_07042005_101813_20668/app_query.fasta_1.1095
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09920705_@CGN_1_1_4087_@runat_07042005_101813_20668 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		4263.5	88.8	3082	3	CNS0A83E	BX820695 Arabidops
c 2		928.5	19.3	767	8	BH552322	BH552322 BOGHQ93TF
3		808.5	16.8	866	9	CL511524	CL511524 SAIL_849_
4		751	15.6	724	8	BH552335	BH552335 BOGHQ93TR
5		747	15.6	812	5	BX840244	BX840244 BX840244
6		716.5	14.9	2706	3	CNS0A2CE	BX842306 Arabidops
7		695.5	14.5	602	4	BG543956	BG543956 E1706 Chi
8		693	14.4	701	2	BE037904	BE037904 AA06B09 A
9		607	12.6	635	1	AU237346	AU237346 AU237346
10		561	11.7	416	8	BH851246	BH851246 SALK_0727
11		537	11.2	582	1	AV829561	AV829561 AV829561
12		535	11.1	817	7	CO095058	CO095058 GR__Ea17I
13		532.5	11.1	800	6	CA766318	CA766318 AF53-Rpf_
14		532.5	11.1	1105	7	CK208950	CK208950 FGAS02067
15		511.5	10.7	582	7	CF322873	CF322873 HDN--02-E
16		510	10.6	685	5	BU831648	BU831648 T024B03 P
17		507	10.6	412	8	BZ763896	BZ763896 SALK_1226
18		500	10.4	281	6	CB260919	CB260919 52-E9517-
19		493	10.3	655	2	AW696401	AW696401 NF103E07S
20		486	10.1	630	4	BG447636	BG447636 NF118B10S
21		470	9.8	295	9	AL938124	AL938124 Arabidops
c 22		461	9.6	362	8	BZ288571	BZ288571 SALK_0219
23		454.5	9.5	624	6	CA032838	CA032838 HX14F14r
24		452	9.4	561	2	BF266387	BF266387 HV_CEa001
25		451	9.4	617	4	BG649121	BG649121 EST510740
26		450.5	9.4	683	7	CK851161	CK851161 11245 Sto
27		450.5	9.4	855	6	CB656009	CB656009 OSJNEc09O
28		448	9.3	662	2	BF641487	BF641487 NF064H03I
c 29		441	9.2	421	5	BP647781	BP647781 BP647781
30		430.5	9.0	441	6	CD481691	CD481691 eca01-17m
31		423	8.8	647	5	BQ696544	BQ696544 NXPV_042_
32		418	8.7	542	7	CO910873	CO910873 BJ03012A1
33		409	8.5	546	7	CN893760	CN893760 010611AAY
34		406.5	8.5	754	6	CA205899	CA205899 SCVPFL107
35		406.5	8.5	793	9	CC634984	CC634984 OGUGZ95TV
36		403	8.4	493	2	BF639297	BF639297 NF030G09I
37		400.5	8.3	814	6	CB633235	CB633235 OSIIEb11P
38		399	8.3	278	9	AL938125	AL938125 Arabidops
39		398	8.3	685	4	BI954416	BI954416 HVSMEm001
c 40		396	8.2	950	9	CL241131	CL241131 ZMMBBb058
41		385	8.0	831	7	CN521077	CN521077 GQ0105.B3
42		384	8.0	477	7	CF759151	CF759151 DSAF1_41_
c 43		379	7.9	854	9	CG960265	CG960265 MBEGS83TR
44		377	7.9	381	5	BQ697581	BQ697581 NXPV_058_
45		376	7.8	666	7	CV092674	CV092674 FAMU_USDA

ALIGNMENTS

RESULT 1
CNS0A83E
LOCUS CNS0A83E 3082 bp mRNA linear HTC 06-FEB-2004
DEFINITION Arabidopsis thaliana Full-length cDNA Complete sequence from clone GSLTPGH60ZD02 of Hormone Treated Callus of strain col-0 of Arabidopsis thaliana (thale cress).
ACCESSION BX820695
VERSION BX820695.1 GI:42467031
KEYWORDS HTC; GSLT_cDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 3082)
AUTHORS Castelli,V., Aury,J.M., Jaillon,O., Wincker,P., Clepet,C., Menard,M., Cruaud,C., Quetier,F., Scarpelli,C., Schachter,V., Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
TITLE Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences: A Combined Approach to Evaluate and Improve Arabidopsis Genome Annotation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3082)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT The sequences are based on single pass reads.
Life Technologies (a division of Invitrogen) members carried out full-length librairies construction : Temple G.
Genoscope members carried out sequencing and annotation : Castelli V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C.,

```
            Schachter V., Weissenbach J., Salanoubat M.
            URGV INRA : Clepet C., Caboche M.
            Annotation is based on the June 2003 version of the Arabidopsis
            genome released by MIPS (Munich Information center for Protein
            Sequences). 5 prime and 3 prime are assembled with Phrap.
            http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
            _length
            http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.
FEATURES              Location/Qualifiers
     source           1. .3082
                      /organism="Arabidopsis thaliana"
                      /mol_type="mRNA"
                      /strain="Col-0"
                      /db_xref="taxon:3702"
                      /clone="GSLTPGH60ZD02"
                      /tissue_type="Hormone Treated Callus"
                      /plasmid="pCMVSPORT_6"
     gene             complement(1. .3082)
                      /gene="At2g43410"
ORIGIN

Alignment Scores:
Pred. No.:              0               Length:         3082
Score:                  4263.50         Matches:        840
Percent Similarity:     92.74%          Conservative:   3
Best Local Similarity:  92.41%          Mismatches:     15
Query Match:            88.80%          Indels:         54
DB:                     3               Gaps:           1

US-09-920-705-3 (1-901) x CNS0A83E (1-3082)

Qy          1 MetAlaLeuSerMetLysPro--PheArgAlaAspAspSerGlyPheGlnSerAsnAsnL 20
              ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db        234 ATGGCGTTATCTATGAAGCCATGTGCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATC 293

Qy         20 euTrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyA 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        294 TTTGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAA 353

Qy         40 rgTyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        354 GATACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACT 413

Qy         60 yrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnG 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        414 ACAGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATG 473

Qy         80 lySerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLysSerLeuTrpValG 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        474 GAAGTCAAATTAAGATCGAATACGCACGACCGGCAAAACCTTGTAAGAGTCTATGGGTGG 533

Qy        100 lyGlyIleGlyProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        534 GTGGAATCGGCCCTAATGTCTCCAAGGATGACCTGGAGGAAGAGTTCAGCAAGTTTGGGA 593

Qy        120 ysIleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluM 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        594 AAATCGAGGATTTTAGGTTTCTCAGAGAACGCAAGACAGCTTTCATTGATTATTATGAGA 653

Qy        140 etAspAspAlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuA 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        654 TGGATGATGCTTTACAGGCTAAGAGCATGAATGGAAAGCCTATGGGTGGTAGCTTTTTGC 713

Qy        160 rgValAspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        714 GTGTTGATTTTCTCCGGTCACAAGCGCCAAAAAAAGAACAATGGGCTGGCTCTTACGATA 773

Qy        180 snArgAsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysG 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        774 ACAGAAATGGCAATATGAATCATAAACCGCAGTATCCTCACTCATATGAAGACTTTAAAG 833

Qy        200 lyAspValGlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysA 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        834 GAGATGTCCAGCCAAGTAAGGTTCTGTGGATTGGGTTCCCTCCTACTGCTACACAATGCA 893

Qy        220 snAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 ATGATGAGCAAATTCTGCACAATGCGATGATACTCTTTGGTGAGATCGAGAGGGTAAAAA 953

Qy        240 erTyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnC 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        954 GTTACCCATCAAGGAATTTTGCACTTGTGGAGTTTAGGAGCGCGGAGGAAGCTCGCCAAT 1013

Qy        260 ysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerA 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1014 GCAAGGAAGGCCTACAGGGGAGGTTATTCAATAATCCTAGAATCAAAATTATGTACTCAA 1073

Qy        280 snAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerA 300
              ||||||||||||||||||||||||||||||||||||||||||
Db       1074 ACGATGAGTTGCCTCCTGAGCAAGACGATACTAGTTTTTACTCT---------------- 1117

Qy        300 rgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIleP 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1117 ---------------------------------------------------------- 1117

Qy        320 roGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrA 340
              |||||||                                                    |||||||
Db       1118 ----------------------------------------------------------GAATACA 1124

Qy        340 snAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1125 ATGACGTTGTTGGTAAGGAGCCAAACTGGAGGAGGCCATCTGCAAATGGAACTGGAATAC 1184

Qy        360 euProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgProM 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1185 TCCCATCTCCAACAGGACCTGGAATCCTCCCATCTCCTGCACAAGGTACGAGGCGCCCTA 1244

Qy        380 etArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuValArgGluSerL 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1245 TGAGGTCAAACCCCGATTCTTGGGAAGGATATGATCCTGCTCAGTTGGTCAGAGAAAGTA 1304

Qy        400 ysArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGluArgS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1305 NACGAACCAGAAGAGATGGATCAGTGGACGGTTTTACTCCNATGGGTGTCGATGAGAGGT 1364

Qy        420 erPheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAspSerAspHisI 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1365 CATTTGGTCGAGGTTCAGTTGCTGCTAGACCTATCCGTGGCCCCCCTGATTCTGATCACA 1424

Qy        440 leTrpArgGlyMetIleAla-LysGlyGlyThrProValCysCysAlaArgCysValPro 459
              |||||||||||||||||||| |||||||||||||||||   ||||||||||||||||||||
Db       1425 TATGGAGAGGAATGATTGCCAAAGGGTGGAACTCCCGTCTGGTGTGCTCGNTGTGTACCT 1484

Qy        460 MetGlyLys-GlyIleGluThrLysLeu-ProGluValValAsnCysSerAlaArgThrA 479
              |||||||| |||:::  |||||||| ||||||||||||||   ||||||   ||||
Db       1485 ATGGGGAAGGGGGATGNAAACTAAACTGCCCTGAGGTCGTCAATTGGTCCGCAGGAACTG 1544

Qy        479 spLeuAsnMetLeuAlaLysHisTyrAlaValAla-IleGlyCysGluIleValPhePhe 498
              |||||||||||||||||||||||||||||||||| :::||||||||||||||||||||||
Db       1545 ATTTGAATATGCTCGCTAAACATTACGCCGTTGCCCATGGGATGTGAGATCGTTTTTTTC 1604

Qy        499 ValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSer 518
              ||||||||||||||||||||||||||||||||||   ||||||||||   ||||||
Db       1605 GTACCAGACAGGGAAGAAGATTTTGCGTCTTACACTGA-TTTCTTCGGTACTTTAGCTCA 1663

Qy        519 LysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProPro 538
              |||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1664 AAAGATCGGGG-GGTGTTGCCAAATTAGATGATGGTACAACTTTATTCTTGGTGCCTCCA 1722

Qy        539 SerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValVal 558
              |||||||   ||||||||||||||||||||||||||||||||||||||||   |||
Db       1723 TCAGATTTCTAAACTGATGTACTCCAAGTGACCCGTCAAGAACGGCTATATGGTGT-GTT 1781

Qy        559 LeuLysLeuProProProAlaValProValThrAlaSerTyrArgGlnGluSerGlnSer 578
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1782 CTCAAGTTACCCCCGCCAGCCGTTCCTGTTACAGCATCATACAGACAAGAATCTCAGTCC 1841
```

```
Qy        579 AsnProLeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeu 598
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1842 AATCCTCTGCATTATATGGATCAAGCCCGGGATTCACCTGCCAATGCTAGTCACAGTTTA 1901

Qy        599 TyrProProArgGluAsnTyrIleArgGlyAlaProGluHisLeuThrAlaAlaSerLys 618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1902 TATCCTCCTAGGGAAAATTACATTAGGGGTGCACCAGAACATTTGACAGCTGCTTCAAAA 1961

Qy        619 ProSerValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSer 638
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1962 CCATCTGTTAGCGAGCCTCTCAGAATACCTAATAATGCAGCGCCTCAAGCTGGGGTTAGT 2021

Qy        639 LeuThrProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAla 658
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2022 TTAACTCCGGAGCTTTTAGCCACTCTGGCATCTATTCTCCCTGCAACTTCTCAACCTGCT 2081

Qy        659 AlaProGluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGln 678
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2082 GCCCCTGAGAGTCACCAACCTATGTCAGGACCTTCAACAGTTGTTTCCACAGCACATCAG 2141

Qy        679 SerAsnGlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThr 698
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2142 TCCAATGGACTGTACAATGGAGAAGCACCGTCTCAAGCTTGGAAAAGAGGTCCACAAACA 2201

Qy        699 ValHisAspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrProAlaGly 718
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2202 GTTCATGATGCGTCAAATCAGTCATTCCAACAATACGGAAATCAGTACACTCCAGCTGGG 2261

Qy        719 GlnLeuProProProProSerArgTyrProProAlaSerAsnAsnProAsnTyrThrSer 738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2262 CAACTACCTCCTCCTCCTTCGCGTTACCCTCCAGCTTCAAACAACCCCAACTACACTAGT 2321

Qy        739 GlyMetValHisGlyAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSer 758
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2322 GGAATGGTCCATGGCAACATGCAATACCAGAGCCAATCTGTTAACATGCCTCAGCTGTCT 2381

Qy        759 ProLeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHis 778
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2382 CCGTTACCAAATATGCCTCATAATAATTATTCCATGTACACTCAGGGTTCGTCAAATCAT 2441

Qy        779 ProValSerGlnProMetValGlnGlnTyrGlnProGluAlaSerMetProAsnGlnAsn 798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2442 CCTGTTTCTCAGCCCATGGTCCAGCAATACCAACCAGAAGCGTCCATGCCAAACCAAAAC 2501

Qy        799 TyrGlyProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAla 818
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2502 TATGGTCCAATTCCAAGTTATCAGCAAGCTAATTTTCATGGCGTAACAACAAATCAGGCA 2561

Qy        819 GlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsn 838
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2562 CAGAACTTAAACCCTTCCCAATTTCAAGCTGCCATGCAACCACCAGCAGATAAGGCAAAT 2621

Qy        839 LeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGly 858
              |||   |||||||||:::|||||||||||||||||||||||||||||||||||   |||
Db       2622 TTATAGCCACAAAACCAATCACTACGATTGCAGCCTATGATCTCTGGGGATGGTCATGGG 2681

Qy        859 --ThrThrAspGlyGluValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaA 878
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2682 TAACAACAGATGGGGAGGTCGATAAGAATCAGAGATACCAGTCAACACTACAATTTGCAG 2741

Qy        878 laAsnLeuLeuLeuGlnIleGlnGlnLysGlnGlnGlnGlnSerSerGlyThrProAlaG 898
              ||||||||||||||||||||||||||||||||||||||||||||    ||||||||||||
Db       2742 CAAACCTTCTTCTCCAGATACAGCAGAAACAGCAGCAACAGTCTTTAGGTACTCCGGCTG 2801

Qy        898 lyGlnGlyPro 901
              ||||||||||
Db       2802 TACAGGGGCCT 2812
```

RESULT 2
BH552322/c

```
LOCUS       BH552322                 767 bp    DNA     linear   GSS 14-DEC-2001
DEFINITION  BOGHQ93TF BOGH Brassica oleracea genomic clone BOGHQ93, genomic
            survey sequence.
ACCESSION   BH552322
VERSION     BH552322.1  GI:17804102
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 767)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGHQ93TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .767
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGHQ93"
                     /clone_lib="BOGH"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

Alignment Scores:
Pred. No.:              3.09e-64       Length:         767
Score:                  928.50         Matches:        198
Percent Similarity:     74.64%         Conservative:   8
Best Local Similarity:  71.74%         Mismatches:     37
Query Match:            19.34%         Indels:         33
DB:                     8              Gaps:           9

US-09-920-705-3 (1-901) x BH552322 (1-767)

Qy        499 ValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSer 518
              :::|||||   |||||||||||||||||||||||||||||||||||||||||    |||
Db        751 CTACCAGACAGCGAAGAAGATTTCGCTTCTTACACTGAGTTTCTCCGCTACCTTGGCTCG 692

Qy        519 LysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProPro 538
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        691 AAAGATCGAGCTGGGGGTGCCAAATTAGACGATGGAACTACTTTATTCCTGGTGCCTCCA 632

Qy        539 SerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValVal 558
              ||||||||||||||||||||||||:::||||||      ||||||||||||||||||||
Db        631 TCAGATTTCTTAACTGATGTACTCAAAGTGACCGGTCCTGAGCGTCTATACGGCGTTGTT 572

Qy        559 LeuLysLeuProProProAlaValProValThrAlaSerTyrArgGlnGluSerGlnSer 578
              |||||||||||||||||||||     ||||||   |||||||||||||||||||||||||
Db        571 CTCAAGCTGCCCCCACCAGCCGCTCCCGTTGCAGCATCGTACAGACAAGAGTCTCAGTCT 512

Qy        579 AsnProLeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeu 598
              ||||||||   |||||||||||   ||||||||||||||||||          ||||||
Db        511 AACCCTCTGTCTTATATGGATCAACCCCGTGATTCGCCTGCCAATACCGGTCACAGCTTT 452

Qy        599 TyrProProArgGluAsnTyrIleArgGlyAlaProGluHisLeuThrAlaAlaSerLys 618
              ||||||||||||                     ||||||||||            |||:::
Db        451 TATCCTCCGAGGGGT---------GCTGCTGCACCAGAACAG------------TCAAGA 413

Qy        619 ProSerValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSer 638
              |||||||||||||||||||||||||:::|||||||||||||||    ||||||||||||||
Db        412 CCATCGGTTAGCGAGCCTCTCAGGTTACCTAACAATGCAGCGTCTCAAGCTGGGGTTAGT 353
```

```
Qy        639 LeuThrProGluLeuLeuAlaThrLeuAlaSerIleLeuPro---AlaThrSerGlnPro 657
              ||||||||||||||||||||||||||||||||||    ||||||   |||:::||||||
Db        352 TTAACCCCGGAGCTTTTAGCTACTCTGGCGTCTTTTCTCCCTGCAGCTTCTTCTCAATCT 293

Qy        658 AlaAlaProGluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHis 677
                |||||||||||||||||||   ||||||      |||:::|||   ||||||
Db        292 ACTGCTCCTGAGAGTCACCAAACTATGTCAGTTGCTTCATCGGTTACTTCCACAGTACCT 233

Qy        678 GlnSerAsnGlyLeuTyrAsnGlyGluAla---ProSerGlnAlaTrpLysArgGlyPro 696
                          ||||||||||||||||      ||||||||||||   |||   |||
Db        232 CAC------------TACAATGGAGAAGCACAGTCTTCTCAAGCTTGGAATAGAGATCCG 185

Qy        697 GlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrPro 716
                                                  |||   ||||||||||||||   |||
Db        184 --------------------------------CAAAATTATGGGAATCAGTACAATCCA 158

Qy        717 AlaGlyGln---LeuProProProProSerArgTyrProProAlaSerAsn------Asn 733
              |||||||||      ||||||||||||   ||||||   ||||||||||||      |||
Db        157 GCTGGACAACCACCTCCTCCTCCTCCGCGTTACGCTCCGGCTTCGAACAACTCCAAT 98

Qy        734 ProAsnTyrThrSerGlyMetValHisGlyAsnMetGlnTyrGlnSerGlnSerValAsn 753
              |||||||||:::||||||||||||||||||||||||||||||||   ||||||||||||
Db         97 CCCAACTACTCCAGTGGAATGGTCCATGGAAACATGCAGTATCAAGGCCAATCTGTTAAC 38

Qy        754 MetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyrSer 769
              |||||||||                |||   ||||||||||:::
Db         37 ATGCCTCAG----------------ATGAATCATAATAATTACACC 8
```

RESULT 3
CL511524

LOCUS	CL511524 866 bp DNA linear GSS 01-APR-2004
DEFINITION	SAIL_849_F10.v1 SAIL Collection Arabidopsis thaliana genomic clone SAIL_849_F10.v1, genomic survey sequence.
ACCESSION	CL511524
VERSION	CL511524.1 GI:46008844
KEYWORDS	GSS.
SOURCE	Arabidopsis thaliana (thale cress)
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE	1 (bases 1 to 866)
AUTHORS	Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D., Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D., Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B., Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
TITLE	A high-throughput Arabidopsis reverse genetics system
JOURNAL	Plant Cell 14 (12), 2985-2994 (2002)
MEDLINE	22356987
PUBMED	12468722
COMMENT	Contact: Sessions A Applied Trait Genetics Syngenta Biotechnology Inc. 3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA Email: allen.sessions@syngenta.com ABRC Stock Number CS837982; T-DNA left border flanking sequences of Syngenta Arabidopsis Insertion Library (SAIL) lines are available through the Arabidopsis Biological Resource Center (ABRC). Sequences represent a pool of amplified genomic regions and not single contiguous sequences. Class: TDNA tagged.

FEATURES	Location/Qualifiers
source	1. .866
	/organism="Arabidopsis thaliana"
	/mol_type="genomic DNA"
	/ecotype="Columbia"
	/db_xref="taxon:3702"
	/clone="SAIL_849_F10.v1"
	/clone_lib="SAIL Collection"
	/note="T-DNA left border sequences were isolated using a modified TAIL-PCR strategy"

ORIGIN

Alignment Scores:

Pred. No.:	1.63e-54	Length:	866
Score:	808.50	Matches:	170
Percent Similarity:	79.82%	Conservative:	12
Best Local Similarity:	74.56%	Mismatches:	32
Query Match:	16.84%	Indels:	14
DB:	9	Gaps:	3

US-09-920-705-3 (1-901) x CL511524 (1-866)

```
Qy        488 AlaValAlaIleGlyCysGluIleValPhePheValProAspArgGluGluAspPheAla 507
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 TCCGTTGCCATTGGATGTGAGATCGTTTTTTTCGTACCAGACAGGGAAGAAGATTTTGCG 242

Qy        508 SerTyrThrGluPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeu 527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 TCTTACACTGAATTTCTCCGGTACCTTAGCTCAAAAGATCGGGCGGGGGTTGCCAAATTA 302

Qy        528 AspAspGlyThrThrLeuPheLeuValProProSerAspPheLeuThrAspValLeuGln 547
              :::||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db        303 GAGGATGGTACAACTTTATTCTTGGTGCCTCCATCACATTTCTTAACTGATGTACTCCAA 362

Qy        548 ValThrArgGlnGluArgLeuTyrGlyValValLeuLysLeuProProProAlaValPro 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 GTGACCCGTCAAGAACGGCTATATGGTGTTGTTCTCAAGTTACCCCCGCCAGCCGTTCCT 422

Qy        568 ValThrAlaSerTyrArgGlnGluSerGlnSerAsnProLeuHisTyrMetAspGlnAla 587
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 GTTACAGTATCATACAGACAAGAATCTCAGTCCAATCCTCTGCATTATATGGATCAAGCC 482

Qy        588 ArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGluAsnTyrIleArg 607
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db        483 CGGGATTCACCTGCCAATGCTAGTCACAGTTTATATCCTCCTAGGGAAAATTACCTTAGG 542

Qy        608 GlyAlaProGluHisLeuThrAlaAlaSerLysProSerValSerGluProLeuArgIle 627
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        543 GGTGCACCAGAACATTTGACAGCTGCTTCAAAACCATCTGTTAGCGAGCCTCTCATAATA 602

Qy        628 ProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAlaThrLeu 647
              ||||||||||||:::||||||||||||||||||||||||||||||||||:::||||||
Db        603 CCTAATAATGCATCGCCTCAAGCTGGGGTTAGTTTAACTCCGGAGCTTTTATCCACTCTG 662

Qy        648 AlaSerIleLeuProAlaThr-SerGlnProAlaAlaProGluSerHisGlnProMetSe 667
              ||||||||||||      ::: |||                  ::::::      |||
Db        663 GCATCTATTCTC------TCTCTCCATTAACACAAGATACCAGGCGTTTTCACCCTGGAA 716

Qy        667 rGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyGluAl 687
                 |||      :::                |||  |||          :::   :::
Db        717 ACCTCCTCGGGCGCTCTATGTNNCCTACAAACAAAGAAAT---------TCGTCCCAGGG 767

Qy        687 aProSerGlnAlaTrpLys----------------------ArgGlyProGlnThrVa 699
              |||:::|||      |||                     |||   |||      ||
Db        768 GCCGGCTCAAGNGGAGAAGCCCCGGTTCTCAGCCTTGGAACAGAGGGCTCCCACACCTGT 827

Qy        699 lHisAspAlaSerAsnGlnSer 706
              ||||   |||         |||
Db        828 TCATTGAGCGGTTAAATCATCC 849
```

RESULT 4
BH552335

LOCUS	BH552335 724 bp DNA linear GSS 14-DEC-2001
DEFINITION	BOGHQ93TR BOGH Brassica oleracea genomic clone BOGHQ93, genomic survey sequence.
ACCESSION	BH552335
VERSION	BH552335.1 GI:17804115
KEYWORDS	GSS.
SOURCE	Brassica oleracea
ORGANISM	Brassica oleracea Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE	1 (bases 1 to 724)
AUTHORS	Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.

```
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGHQ93TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .724
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGHQ93"
                     /clone_lib="BOGH"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

Alignment Scores:
Pred. No.:              5.22e-50        Length:         724
Score:                  751.00          Matches:        150
Percent Similarity:     70.29%          Conservative:   18
Best Local Similarity:  62.76%          Mismatches:     24
Query Match:            15.64%          Indels:         47
DB:                     8               Gaps:           4

US-09-920-705-3 (1-901) x BH552335 (1-724)

Qy        173 GlnTrpAlaGlySerTyrAspAsnArgAsnGlyAsn---MetAsnHisLysProGln--- 190
              :::|||||||||||||||||||||||||||||||   ||||||||||||||||
Db         55 AAATGGGCTGGCTCTTACGATAACAGGAATGGCAATGTGATGAATCACAAACCTCAGGTT 114

Qy        190 ------------------------------------------------------------ 190

Db        115 CGTTTTAAAATATTCAAAAAAGAGTTTGTTTATTGTTTCTGCTTATATGTGTATGATGTT 174

Qy        191 ----------------------------TyrProHisSerTyrGluAspPheLysGlyAs 201
                                           ||||||||||||:::||||||   :::|||||
Db        175 AATTAGCTTTTTGACGTTTTGTTGTTAGTATCCTCACTCACATGAAGACGCCAGAGGAGA 234

Qy        201 pValGlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAs 221
              |   |||||||||||||||||||||||||:::|||||:::                 ||
Db        235 TGACCAGCCAAGTAAAGTTCTGTGGATTGGGTACCCTCCTTCTGTTCAG------ATAGA 288

Qy        221 pGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTy 241
              |||||||:::|||||||||||||||||||||||||||||||||||||   ||||||||
Db        289 CGAGCAAATGCTACACAACGCAATGATACTCTTCGGTGAGATCGAGAGGAAGAAAAGTTA 348

Qy        241 rProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLy 261
              ||||||||||:::|||:::|||||||||||||||||||   :::||||||||||||  ||
Db        349 CCCGTCGAGGCATTTTTCACTTGTGGAGTTTAGGAGCGTGGACGAAGCTCGCCAAGCCAA 408

Qy        261 sGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAs 281
              |||||||||||||||||||||||   :::|||||||||   ||||||||||||||||||
Db        409 GGAAGGGCTACAAGGGAGGCTATTCAAGGATCCTAGAATCACTATCATGTACTCGAACGA 468

Qy        281 pGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArgTh 301
              ||||:::|||||||||||||||||||   ||||||||||||||:::||||||||||||
Db        469 TGAGATACCTCCTGAGCAAGATGATAATAGTTTTTATTCTGGCGTGAAACGGTCAAGGCC 528

Qy        301 rAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIleProGl 321
                  ||||||      ||||||
Db        529 AGGTATGTTCATCACTGATCCT---------------------------------------- 550

Qy        321 ySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAs 341
               |||::::::|||   ||||||:::|||:::||||||||||||   ::::::|||||||
Db        551 -TCATTGAAGCCATTCAGAGGTAGCAATGACCGTTCTTATAATGTTTCAGATTACAATGA 609

Qy        341 pValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeuPr 361
              ||||||||||   ||||||||||||||||||||||||   ||||||||||||||||||
Db        610 CGTTGTTGGTATGGAGCCAAACTGGAGGAGGCCGTCTCCAAATGGGACTGGGATACTCCC 669

Qy        361 oSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgPro 379
              |||||||   |||   ||||||   ||||||||||||||||||   |||   |||
Db        670 ATCTCCAGCAGGACATGGGATCCCTCCGTCTCCTGCACAAGGGATGAGGAACCCT 724

RESULT 5
BX840244
LOCUS       BX840244                812 bp    mRNA    linear   EST 11-FEB-2004
DEFINITION  BX840244 Arabidopsis thaliana Hormone Treated Callus Col-0
            Arabidopsis thaliana cDNA clone GSLTPGH87ZF04 5PRIM, mRNA sequence.
ACCESSION   BX840244
VERSION     BX840244.1  GI:42534327
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 812)
  AUTHORS   Castelli,V., Aury,J.M., Jaillon,O., Wincker,P., Clepet,C.,
            Menard,M., Cruaud,C., Quetier,F., Scarpelli,C., Schachter,V.,
            Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
  TITLE     Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:
            A Combined Approach to Evaluate and Improve Arabidopsis Genome
            Annotation
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            The sequences are based on single pass reads.
            Life Technologies (a division of Invitrogen) members carried out
            full-length librairies construction : Temple G.
            Genoscope members carried out sequencing and annotation : Castelli
            V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C.,
            Schachter V., Weissenbach J., Salanoubat M.
            URGV INRA : Clepet C., Caboche M.
            Annotation is based on the June 2003 version of the Arabidopsis
            genome released by MIPS (Munich Information center for Protein
            Sequences).
            http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/EST
            http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.
FEATURES             Location/Qualifiers
     source          1. .812
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /ecotype="Col-0"
                     /db_xref="taxon:3702"
                     /clone="GSLTPGH87ZF04"
                     /tissue_type="Hormone Treated Callus"
                     /clone_lib="Arabidopsis thaliana Hormone Treated Callus
                     Col-0"
ORIGIN

Alignment Scores:
Pred. No.:              1.3e-49         Length:         812
Score:                  747.00          Matches:        148
Percent Similarity:     86.84%          Conservative:   17
Best Local Similarity:  77.89%          Mismatches:     23
Query Match:            15.56%          Indels:         2
DB:                     5               Gaps:           0

US-09-920-705-3 (1-901) x BX840244 (1-812)

Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 305

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
```

```
Db        306 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGCAGA 365

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60

Db        366 TAAGGCGAAATTGATAGACTCCCGGTGTATTCTTCACGAGGCTTTGCGTTTATCTCCTAC 425

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80

Db        426 AGACATGTGGAGGCAGCAGTCGCAGCCCAAGAGGCTCTTCAAGGAGCAAATTTGAATGGC 485

Qy         81 SerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGly 100

Db        486 AGTCAAATTCAGATAGAATACGCAAGACCGGCCAAGCCTTGTCAGAGTCTATGGGTGGGT 545

Qy        101 GlyIleGlyProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLys 120

Db        546 GGCATCGGCACTAATGTATCCAAAGATGAACAGGAAGCAAAGTTCAGCAAGTTTGGGAAA 605

Qy        121 IleGluAspPhe-ArgPheLeuArgGluArgLysThrAlaPheIleAspTyr-TyrGluM 140

Db        606 GTCAAGGATTTTACGGGTTTCTCAGAAAACGAAGACAGATTTCCTTGATTATTCATGAGA 665

Qy        140 etAspAspAlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuA 160

Db        666 TGGATGATGATTTACCAGATAAGAGCATGACTGGGAAGCCACTGGGTGGTCGCTTTTTGC 725

Qy        160 rgValAspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspA 180

Db        726 GTGTTAAATTTCTCCGGACACCAAAGACAAAGGGAGAAAAATGGGCTGACTCTTACGATA 785

Qy        180 snArgAsnGlyAsnMetAsnHisLys 188

Db        786 ACAGAAATGGAAATTAGAATAATAAA 811
```

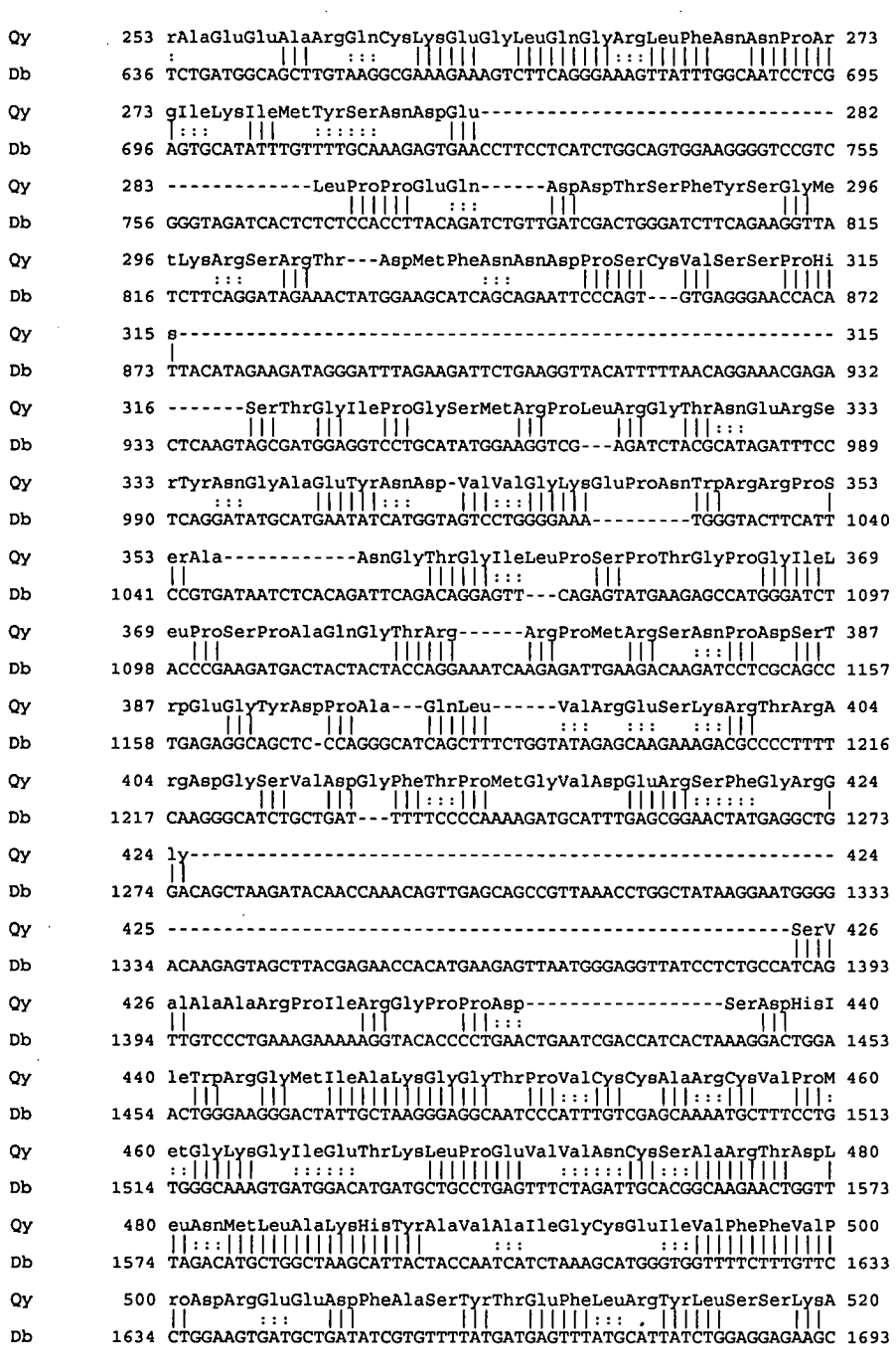
```
RESULT 6
CNS0A2CE
LOCUS       CNS0A2CE                2706 bp    mRNA    linear   HTC 04-FEB-2004
DEFINITION  Arabidopsis thaliana Full-length cDNA Complete sequence from clone
            GSLTFB49ZD02 of Flowers and buds of strain col-0 of Arabidopsis
            thaliana (thale cress).
ACCESSION   BX842306
VERSION     BX842306.1  GI:42406938
KEYWORDS    HTC; GSLT_cDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2706)
  AUTHORS   Castelli,V., Aury,J.M., Jaillon,O., Wincker,P., Clepet,C.,
            Menard,M., Cruaud,C., Quetier,F., Scarpelli,C., Schachter,V.,
            Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
  TITLE     Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:
            A Combined Approach to Evaluate and Improve Arabidopsis Genome
            Annotation
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2706)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     The sequences are based on single pass reads.
            Life Technologies (a division of Invitrogen) members carried out
            full-length librairies construction : Temple G.
            Genoscope members carried out sequencing and annotation : Castelli
            V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C.,
            Schachter V., Weissenbach J., Salanoubat M.
            URGV INRA : Clepet C., Caboche M.
            Annotation is based on the June 2003 version of the Arabidopsis
            genome released by MIPS (Munich Information center for Protein
            Sequences). 5 prime and 3 prime are assembled with Phrap.
            http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
            _length
            http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.
FEATURES              Location/Qualifiers
     source           1. .2706
                      /organism="Arabidopsis thaliana"
                      /mol_type="mRNA"
                      /strain="Col-0"
                      /db_xref="taxon:3702"
                      /clone="GSLTFB49ZD02"
                      /tissue_type="Flowers and buds"
                      /plasmid="pCMVSPORT_6"
     gene             1. .2706
                      /gene="At4g12630-At4g12640"
ORIGIN
```

Alignment Scores:

Pred. No.:	2.16e-46	Length:	2706
Score:	716.50	Matches:	271
Percent Similarity:	38.73%	Conservative:	112
Best Local Similarity:	27.40%	Mismatches:	298
Query Match:	14.92%	Indels:	310
DB:	3	Gaps:	37

US-09-920-705-3 (1-901) x CNS0A2CE (1-2706)

```
Qy         17 SerAsnAsnLeuTrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGlu 36

Db        108 TCGCGACATCTTTGGGTTGGTAACCTTCCTCATGGGATACTTGAACGAGAATTAGCCGAT 167

Qy         37 LeuPheGlyArgTyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAla 56

Db        168 CGGTTCTTAAGATTTGGTGAATTGGAAAGCTTAGCGTTTCAGCCTGGACGAAGCTACGCA 227

Qy         57 PheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAla 76

Db        228 TTTGTTAATTTTAATCACGATGAGGATGCGTTCGCTGCGATTGAGTCGCTTCAAGGCTTC 287

Qy         77 AsnLeuAsnGlySerGlnIleLysIle-GluTyrAlaArgProAlaLysProCysLysSe 96

Db        288 CCACTCTCTGGAAACCCACTTAGGATCGGAGTTTGCTAAGGCGGGGAAG---------- 336

Qy         96 rLeuTrpValGlyGlyIleGlyProAsnValSerLysAspAspLeuGluGluGluPheSe 116

Db        337 ----------------------------------------------------------TC 338

Qy        116 rLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIleAs 136

Db        339 ATCAACTGGATCACGCACTGAT----------------------------------GA 362

Qy        136 pTyrTyrGluMetAspAspAlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGl 156

Db        363 TATATATCGCCATGAA--------------AAACAACGGTCTAAAAAAAAAAAAGGGTC 407

Qy        156 ySerPheLeuArgValAspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGl 176

Db        408 GTCTTTTGTCCAA---------------------------------------------- 420

Qy        176 ySerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGl 196

Db        421 -------------AGGGACTCCAGAATGCGTTAT------GAGAGCCCAGACACATATAG 461

Qy        196 uAspPheLys---------GlyAspValGlnProSerLysValLeuTrpIleGlyPhePr 213

Db        462 CAAATCAAAAATGAATGATAGAAATGCAGAACCCAGTGAGGTACTATATATAGGGTTTCC 521

Qy        213 oProThrAlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGl 233

Db        522 T------GCTTCGTTGAAAGTAGATGATGCGCTCTTGAGGAACGTCTTTTCTTCGTTTGG 575

Qy        233 yGluIleGluArgValLysSerTyrProSerArgAsnPheAlaLeuValGluPheArgSe 253

Db        576 AGAAATAACAAAGGTCACGGTATTCCCTGGTCGCAGTTATGCATTTGTTCAATTCCGAAA 635
```


Qy 253 rAlaGluGluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProAr 273
Db 636 TCTGATGGCAGCTTGTAAGGCGAAAGAAAGTCTTCAGGGAAAGTTATTTGGCAATCCTCG 695
Qy 273 gIleLysIleMetTyrSerAsnAspGlu---------- 282
Db 696 AGTGCATATTTGTTTTGCAAAGAGTGAACCTTCCTCATCTGGCAGTGGAAGGGGTCCGTC 755
Qy 283 ----------LeuProProGluGln------AspAspThrSerPheTyrSerGlyMe 296
Db 756 GGGTAGATCACTCTCTCCACCTTACAGATCTGTTGATCGACTGGGATCTTCAGAAGGTTA 815
Qy 296 tLysArgSerArgThr---AspMetPheAsnAsnAspProSerCysValSerSerProHi 315
Db 816 TCTTCAGGATAGAAACTATGGAAGCATCAGCAGAATTCCCAGT---GTGAGGGAACCACA 872
Qy 315 s---------- 315
Db 873 TTACATAGAAGATAGGGATTTAGAAGATTCTGAAGGTTACATTTTTAACAGGAAACGAGA 932
Qy 316 -------SerThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSe 333
Db 933 CTCAAGTAGCGATGGAGGTCCTGCATATGGAAGGTCG---AGATCTACGCATAGATTTCC 989
Qy 333 rTyrAsnGlyAlaGluTyrAsnAsp-ValValGlyLysGluProAsnTrpArgArgProS 353
Db 990 TCAGGATATGCATGAATATCATGGTAGTCCTGGGGAAA---------TGGGTACTTCATT 1040
Qy 353 erAla----------AsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleL 369
Db 1041 CCGTGATAATCTCACAGATTCAGACAGGAGTT---CAGAGTATGAAGAGCCATGGGATCT 1097
Qy 369 euProSerProAlaGlnGlyThrArg------ArgProMetArgSerAsnProAspSerT 387
Db 1098 ACCCGAAGATGACTACTACTACCAGGAAATCAAGAGATTGAAGACAAGATCCTCGCAGCC 1157
Qy 387 rpGluGlyTyrAspProAla---GlnLeu------ValArgGluSerLysArgThrArgA 404
Db 1158 TGAGAGGCAGCTC-CCAGGGCATCAGCTTTCTGGTATAGAGCAAGAAAGACGCCCCTTTT 1216
Qy 404 rgAspGlySerValAspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgG 424
Db 1217 CAAGGGCATCTGCTGAT---TTTTCCCCAAAAGATGCATTTGAGCGGAACTATGAGGCTG 1273
Qy 424 ly---------- 424
Db 1274 GACAGCTAAGATACAACCAAACAGTTGAGCAGCCGTTAAACCTGGCTATAAGGAATGGGG 1333
Qy 425 ----------SerV 426
Db 1334 ACAAGAGTAGCTTACGAGAACCACATGAAGAGTTAATGGGAGGTTATCCTCTGCCATCAG 1393
Qy 426 alAlaAlaArgProIleArgGlyProProAsp----------SerAspHisI 440
Db 1394 TTGTCCCTGAAAGAAAAAGGTACACCCCTGAACTGAATCGACCATCACTAAAGGACTGGA 1453
Qy 440 leTrpArgGlyMetIleAlaLysGlyGlyThrProValCysCysAlaArgCysValProM 460
Db 1454 ACTGGGAAGGGACTATTGCTAAGGGAGGCAATCCCATTTGTCGAGCAAAATGCTTTCCTG 1513
Qy 460 etGlyLysGlyIleGluThrLysLeuProGluValValAsnCysSerAlaArgThrAspL 480
Db 1514 TGGGCAAAGTGATGGACATGATGCTGCCTGAGTTTCTAGATTGCACGGCAAGAACTGGTT 1573
Qy 480 euAsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGluIleValPhePheValP 500
Db 1574 TAGACATGCTGGCTAAGCATTACTACCAATCATCTAAAGCATGGGTGGTTTTCTTTGTTC 1633
Qy 500 roAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysA 520
Db 1634 CTGGAAGTGATGCTGATATCGTGTTTTATGATGAGTTTATGCATTATCTGGAGGAGAAGC 1693
Qy 520 spArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProProSerA 540
Db 1694 AACGGGCAGCTGTTTCTAAATTGGATGACACAACAACGTTGTTTCTGGTGCCTCCATCTG 1753
Qy 540 spPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuL 560
Db 1754 ATTTCTCCGAGAAAGTACTTAAAGTTCCTGGGAAACTAAGCATCTCTGGAGTTATTCTAC 1813
Qy 560 ysLeuProProProAlaValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnP 580
Db 1814 GTTTA---------- 1818
Qy 580 roLeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrP 600
Db 1819 ----------GAATGTGGTGGTTCTGGCTCCGGTGGGC 1846
Qy 600 roProArgGluAsnTyrIleArgGlyAlaProGluHisLeuThrAlaAlaSerLysProS 620
Db 1847 CTGTTCAACAGCAAGGTGAGAGAAAAGATACAGACTTGCTAACTTATTATGGTGAAACAT 1906
Qy 620 erValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuT 640
Db 1907 CATATTCAGAA---------- 1917
Qy 640 hrProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaP 660
Db 1918 ----------GCTAGTGGAGCGTTTCCTGA-TGTGGAAACCCACGTATTC 1956
Qy 660 roGluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerA 680
Db 1957 CG----------GGTCCAACAGCTTTTCTGCGTAGTGCGGGTCGTGACA 1995
Qy 680 snGlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValH 700
Db 1996 AT----------CAAAGTGCATCAATGGATCCATAC---GTCG 2025
Qy 700 isAspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGlnL 720
Db 2026 AGGACAAGCATGATCAGTTGAGTCACCGGTATTCTGGA---------- 2063
Qy 720 euProProProProSerArgTyrProProAlaSerAsnAsnProAsnTyrThr------- 737
Db 2064 ----------TCGGACTGGCCGCCTCGTGATACAAACCCCAGAAGTTCACCCTTCA 2109
Qy 738 ----------SerGlyMetValHisGlyAsnMetGlnTyrGlnSerG 750
Db 2110 TAGATCATACTGTTCAAAAACACAGCGGGTTTGTTCCTGGGAAACAGCAA---------- 2159
Qy 750 lnSerValAsnMetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyrSerM 770
Db 2160 ----------AATGCAGATCTAAGTC 2175
Qy 770 etTyrThrGlnGlySerSerAsnHisProValSerGlnProMetValGlnGlnTyrGlnP 790
Db 2176 GTTAT----------CATGATACGGAAACTCCAGTCCCAGCAGGGTTTCAGC 2217
Qy 790 roGlu----------AlaSerMetProAsnGlnAsnTyrGlyProIleProS 804
Db 2218 CTGAACAGCTAACACATTTAGCTCCCTCGCTACCCAGACAG----------C 2259
Qy 804 erTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsn---------L 821
Db 2260 AACAACAGGAACAGAAC----------ACGCCAAATCAGCCTGAGAGATTTTTTTTT 2307
Qy 821 euAsnProSerGlnPheGln------AlaAlaMetGlnProProAlaAspLysAlaAsnL 839
Db 2308 TAAGGCCGAGTTTGTTTCAGCCATTTGCAACATGAGCAGACGCCAAGCATACCTCAACCG 2367
Qy 839 euGluProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyT 859
Db 2368 TAACTCCACAAAATCAAAATGTACAGATACAA---------AGTAGCAACAGCCAACAAC 2418
Qy 859 hrThrAspGlyGluValAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaA 879

```
                 :::  |||  :::      :::|||  |||:::|||||||||  ||||||
Db       2419 AAGAAGAAACAGAGGCCAATCCACAGAAGCGTCTTCAAGCAACATTACAGCTAGCAGCTG 2478

Qy        879 snLeuLeuLeuGlnIleGlnGln 886
              ||||||  ||||||||||||
Db       2479 GACTTCTCCAGCAGATTCAACAA 2501
```

RESULT 7
BG543956

```
LOCUS       BG543956                602 bp    mRNA    linear   EST 01-MAY-2002
DEFINITION  E1706 Chinese cabbage etiolated seedling library Brassica rapa
            subsp. pekinensis cDNA clone E1706, mRNA sequence.
ACCESSION   BG543956
VERSION     BG543956.1  GI:20374936
KEYWORDS    EST.
SOURCE      Brassica rapa subsp. pekinensis
  ORGANISM  Brassica rapa subsp. pekinensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   Ryu,S.H., Kang,J.S., Kang,C.-h., Kim,C.Y., Choi,Y.J., Lee,S.-H.,
            Bahk,J.D., Lee,S.Y., Cho,M.J. and Lim,C.O.
  TITLE     Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Lim, C.O.
            Plant Molecular Biology & Biotechnology Research Centre
            Gyeongsang National University
            #900 Gazwa-dong, Jinju 660-701, Korea
            Tel: 82 55 751 6255
            Fax: 82 55 759 9363
            Email: colim@nongae.gsnu.ac.kr
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .602
                     /organism="Brassica rapa subsp. pekinensis"
                     /mol_type="mRNA"
                     /cultivar="Jangwon"
                     /sub_species="pekinensis"
                     /db_xref="taxon:51351"
                     /clone="E1706"
                     /tissue_type="Etiolated seedling"
                     /lab_host="XL-1 Blue"
                     /clone_lib="Chinese cabbage etiolated seedling library"
                     /note="Vector: pSPORT 1; Site_1: Sal I; Site_2: Not I"
ORIGIN
```

Alignment Scores:

Pred. No.:	1.15e-45	Length:	602
Score:	695.50	Matches:	138
Percent Similarity:	84.02%	Conservative:	25
Best Local Similarity:	71.13%	Mismatches:	26
Query Match:	14.49%	Indels:	6
DB:	4	Gaps:	2

US-09-920-705-3 (1-901) x BG543956 (1-602)

```
Qy         23 GlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArgTyrGly 42
              ||||||||||||||:::|||||||||||||||||||   :::|||||||||||:::|||
Db         13 GGCAGTCTGACCCCCGATACTACCGAATCCGATCTCGCCGATCTGTTCGGAAGGTTCGGC 72

Qy         43 AspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyrArgHis 62
              ||||||||||:::|||   ||||||||||   ||||||||||||||:::|||
Db         73 GATATCGACAGGCTCACCGCCTATTCTTCCCGTTGCTTCGCGTTTATCTACTTCAGGCGC 132

Qy         63 ValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGln 82
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        133 GTTGAGGAAGCCGTAGCCGCGAAGGAGGCTCTTCAGGGTGCCAATCTAAACGGGAGTCGC 192

Qy         83 IleLysIleGluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGlyGlyIle 102
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        193 ATTAAAATCGAATTCGCACGACCGGCAAAACCTTGTAAGAGTCTTTGGGTTGGTGGAATC 252

Qy        103 GlyProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGlu 122
                 :::|||  ||||||  ||||||  ||||||  |||||||||:::||||||
Db        253 AGCTCGAGTGTCCCGAAGGATGTACTCGAGGCAGAGTTCAGGAAGTTTGGGGAAATCGAA 312

Qy        123 AspPheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAsp 142
              ||||||||||||:::::|||||||||:::||||||||||||:::::::::||||||
Db        313 AGTTTCAGGTTTCTTAAGGATCGCAAGACGTCTTTTATTGACTATTTTGACGTGGACGAT 372

Qy        143 AlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAsp 162
              |||:::||||||||:::|||||||||||||   :::|||||||||||||||||||||||
Db        373 GCTGTACAGGCAAAGACCATGAACGGGAAGCGACTAGGTGGTAGCTTTTTGCGAGTTGAT 432

Qy        163 PheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsn 182
              ||||||||||||||   |||:::||||||   ||||||   |||   |||:::
Db        433 TTTCTTCGCTCACAAGGGCCAAGAAAAGAACCACGGCTTGGCTCCGAAGATGGCAGGGAT 492

Qy        183 GlyAsnMetAsnHisLys---ProGlnTyrProHisSerTyrGluAspPheLysGlyAsp 201
              |||::::::::::   |||   ||||||   ||||||                |||
Db        493 GGCAGTTTGAGTGCTAAGCACCAGCAGTATACTCACTCA-----------CCTAGAGAT 540

Qy        202 ValGlnProSerLysValLeuTrpIleGlyPheProProThr 215
              |||   |||  ||||||||||||||:::||||||:::
Db        541 GGCCAACA-AGCAACGTTCTGTGGATTGGGTACCCTCCTTCT 581
```

RESULT 8
BE037904

```
LOCUS       BE037904                701 bp    mRNA    linear   EST 07-JUN-2000
DEFINITION  AA06B09 AA Arabidopsis thaliana cDNA 5', mRNA sequence.
ACCESSION   BE037904
VERSION     BE037904.1  GI:8332920
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 701)
  AUTHORS   Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M.,
            Ferrea,H., Kawasaki,S., McCollough,A., Michalowski,C.B.,
            Palacio,C., Scara,G., Wheeler,M. and Zepeda,G.R.
  TITLE     Functional Genomics of Plant Stress Tolerance
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Michalowski,C.B.
            University of Arizona
            Bio Sciences West room 513, Tucson, AZ 85721, USA
            Tel: 520-621-7982
            Fax: 520-621-1697
            Email: cbm@u.arizona.edu
            Best blastx match: 'gb|AAB64314.1| (AC002335) hypothetical protein
            [Arabidopsis thal. . . 165 3e-40'. An open reading frame exists.
            Insert Length: 2   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /tissue_type="leaves, flowering plants"
                     /dev_stage="12 weeks"
                     /clone_lib="AA"
                     /note="20 h 200mM NaCl"
ORIGIN
```

Alignment Scores:

Pred. No.:	2.28e-45	Length:	701
Score:	693.00	Matches:	140
Percent Similarity:	96.62%	Conservative:	3
Best Local Similarity:	94.59%	Mismatches:	5
Query Match:	14.43%	Indels:	2
DB:	2	Gaps:	0

US-09-920-705-3 (1-901) x BE037904 (1-701)

```
Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 304

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 364

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 424

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 484

Qy         81 SerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGly 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 AGTCAAATTAAGATCGAATACGCACGACCGGCAAAACCTTGTAAGAGTCTATGGGTGGGT 544

Qy        101 GlyIleGlyProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLys 120
              ||||||||||||||||||||||||||||||:::|||||||||||||||   ||||||
Db        545 GGAATCGGTCCTAATGTTTCCAAGGATGACCTGCAGGAAGAGTTCAGCAAGCTTGGGAAA 604

Qy        121 IleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMet 140
              ||||||||||||:::   ||||||||||||||||||||||||||   ||||||||||||
Db        605 ATCGAGGATTTTAAGGTCCTCAGAGAACGCAAGACAGCTTTCATCGA-TATTATGAGATG 663

Qy        141 AspAspAlaLeuGlnAlaLysSer 148
                |||   ||||||||||:::
Db        664 GCTGATGC-TTACAGGCTAAGACA 686

RESULT 9
AU237346
LOCUS       AU237346                635 bp    mRNA    linear   EST 01-APR-2002
DEFINITION  AU237346 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-16-A15 5',
            mRNA sequence.
ACCESSION   AU237346
VERSION     AU237346.1  GI:19876515
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Akiyama,K., Enju,A., Oono,Y., Sakurai,T., Carninci,P., Kawai,J.,
            Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A.,
            Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998).cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further
            details.
FEATURES             Location/Qualifiers
     source          1. .635
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL16-16-A15"
                     /lab_host="DH10B"
                     /clone_lib="RAFL16"
                     /note="Site_1: BamHI; Site_2: SalI; dark-grown"
ORIGIN

Alignment Scores:
Pred. No.:              1.61e-38        Length:         635
Score:                  607.00          Matches:        123
Percent Similarity:     97.62%          Conservative:   0
Best Local Similarity:  97.62%          Mismatches:     2
Query Match:            12.64%          Indels:         2
DB:                     1               Gaps:           0

US-09-920-705-3 (1-901) x AU237346 (1-635)

Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 314

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 374

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 434

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 494

Qy         81 SerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLysSerLeuTrp-ValGl 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        495 AGTCAAATTAAGATCGAATACGCACGACCGGCAAAACCTTGTAAGAGTCTATGGGGTGGG 554

Qy        100 yGlyIleGlyProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLy 120
              |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        555 TGGAATCGGGCCTAATGTCTCCAAGGATGGGCTGGAGGAAGAGTTCAGCAAGTTTGGNAA 614

Qy        120 sIleGluAspPheArg 125
              |   ||||||||||||
Db        615 ATC-GAGGATTTTAGG 629

RESULT 10
BH851246
LOCUS       BH851246                416 bp    DNA     linear   GSS 13-JUN-2002
DEFINITION  SALK_072730.56.00.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_072730.56.00.x, genomic
            survey sequence.
ACCESSION   BH851246
VERSION     BH851246.1  GI:21422117
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
            Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
            Shinn,P., Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA. This sequence lies within an annotated exon of At2g43410.
            Class: TDNA tagged.
```

```
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Col-0"
                     /db_xref="taxon:3702"
                     /clone="SALK_072730.56.00.x"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
ORIGIN

Alignment Scores:
Pred. No.:                4.3e-35        Length:          416
Score:                    561.00         Matches:         107
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              11.69%         Indels:          0
DB:                       8              Gaps:            0

US-09-920-705-3 (1-901) x BH851246 (1-416)

Qy        795 ProAsnGlnAsnTyrGlyProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThr 814
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 CCAAACCAAAACTATGGTCCAATTCCAAGTTATCAGCAAGCTAATTTTCATGGCGTAACA 61

Qy        815 ThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAla 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 ACAAATCAGGCACAGAACTTAAACCCTTCCCAATTTCAAGCTGCCATGCAACCACCAGCA 121

Qy        835 AspLysAlaAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGly 854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 GATAAGGCAAATTTAGAGCCACAAAACCAAGCACTACGATTGCAGCCTATGATCTCTGGG 181

Qy        855 AspGlyGlnGlyThrThrAspGlyGluValAspLysAsnGlnArgTyrGlnSerThrLeu 874
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GATGGTCAGGGTACAACGGATGGGGAGGTCGATAAGAATCAGAGATACCAGTCAACACTA 241

Qy        875 GlnPheAlaAlaAsnLeuLeuLeuGlnIleGlnGlnLysGlnGlnGlnGlnSerSerGly 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 CAATTTGCAGCAAACCTTCTTCTCCAGATACAGCAGAAACAGCAGCAACAGTCTTCAGGT 301

Qy        895 ThrProAlaGlyGlnGlyPro 901
              |||||||||||||||||||||
Db        302 ACTCCGGCTGGACAGGGGCCT 322

RESULT 11
AV829561
LOCUS       AV829561                582 bp    mRNA    linear   EST 01-APR-2002
DEFINITION  AV829561 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-47-P10 5',
            mRNA sequence.
ACCESSION   AV829561
VERSION     AV829561.1  GI:19871621
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998).cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further
            details.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL09-47-P10"
                     /dev_stage="plants at various developmental stages from
                     germination to mature seeds"
                     /lab_host="DH10B"
                     /clone_lib="RAFL9"
                     /note="Site_1: BamHI; Site_2: SalI; subjected to
                     dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
                     hr) treatments"
ORIGIN

Alignment Scores:
Pred. No.:                5.99e-33       Length:          582
Score:                    537.00         Matches:         108
Percent Similarity:       97.30%         Conservative:    0
Best Local Similarity:    97.30%         Mismatches:      1
Query Match:              11.19%         Indels:          2
DB:                       1              Gaps:            0

US-09-920-705-3 (1-901) x AV829561 (1-582)

Qy          1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 ATGGCGTTATCTATGAAGCCATTCAGAGCCGATGATTCCGGTTTCCAGTCAAACAATCTT 313

Qy         21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 TGGGTCGGTAGCCTAACGCCGGAGACGACAGAGTCAGATCTGACCGAGTTGTTTGGAAGA 373

Qy         41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        374 TACGGCGATATTGATAGAATCACGGTGTATTCTTCACGAGGCTTTGCGTTTATATACTAC 433

Qy         61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        434 AGACATGTGGAGGAAGCAGTCGCAGCCAAAGAGGCTCTTCAAGGAGCAAATTTGAATGGA 493

Qy         81 SerGlnIleLysIleGluTyrAlaArgProAlaLys-ProCysLysSerLeuTrp-ValG 100
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||| ||||
Db        494 AGTCAAATTAAGATCGAATACGCACGACCGGCAAAACCCTTGTAAGAGTCTATGGGGTGG 553

Qy        100 lyGlyIleGlyProAsnValSerLysAsp 109
              ||||||||   ||||||||||||||||||
Db        554 GTGGAATCNGCCCTAATGTNTCCAAGGAT 582

RESULT 12
CO095058
LOCUS       CO095058                817 bp    mRNA    linear   EST 16-JUN-2004
DEFINITION  GR__Ea17I03.r GR__Ea Gossypium raimondii cDNA clone GR__Ea17I03 3',
            mRNA sequence.
ACCESSION   CO095058
VERSION     CO095058.1  GI:48793744
KEYWORDS    EST.
SOURCE      Gossypium raimondii
  ORGANISM  Gossypium raimondii
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
```

```
REFERENCE   1  (bases 1 to 817)
  AUTHORS   Kim,H., Yu,Y., Kudrna,D., Hatfield,J., Stum,D., Mueller,C.,
            Udall,J.A., Rapp,R.A., Wendel,J.F., Rao,K., Soderlund,C. and
            Wing,R.A.
  TITLE     Global assembly of Cotton ESTs
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            The University of Arizona
            Forbes Building Room 303, Tucson, AZ, 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 17  row: I  column: 03.
FEATURES             Location/Qualifiers
     source          1. .817
                     /organism="Gossypium raimondii"
                     /mol_type="mRNA"
                     /db_xref="taxon:29730"
                     /clone="GR__Ea17I03"
                     /tissue_type="whole seedlings"
                     /dev_stage="first true leaves"
                     /lab_host="DH10B"
                     /clone_lib="GR__Ea"
                     /note="Vector: pCMV.SPORT-6.1; Site_1: NotI; Site_2:
                     EcoRV; Library made by Invitrogen with RNA supplied by
                     Wendle lab. Directional cloned into NotI-EV.  Colonies
                     plated/picked by AGI. More glycerol clones held in -80."
ORIGIN

Alignment Scores:
Pred. No.:                1.43e-32        Length:         817
Score:                    535.00          Matches:        133
Percent Similarity:       57.60%          Conservative:   30
Best Local Similarity:    47.00%          Mismatches:     64
Query Match:              11.14%          Indels:         57
DB:                       7               Gaps:           9

US-09-920-705-3 (1-901) x CO095058 (1-817)

Qy        292 PheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsnAsp---------- 307
              |||   ||||||:::        |||      |||         |||
Db         16 TTTGATTCAGGCATACATGGAAAAAGGCCTGGCATGCTGTTTGGTGACCATCTATTTAGG 75

Qy        308 ----------------------------------------ProSerCysValSer--- 312
                                                     |||:::   ||||||
Db         76 CCTTCACAAATGGACATTCTTGGTCAAAATAATCCATTACTACCAAATACGGTCTCTGGA 135

Qy        313 SerProHisSerThrGlyIleProGlySerMetArgProLeuArg-------------- 327
              |||      |||:::||||||   ||||||      |||:::|||
Db        136 TCATTACCATCTAGTGGCATTTTTGGATCAAATGTGCCGATCAGGTCTTTTAGTCATCAA 195

Qy        328 GlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValValGlyLysGluPro 347
              |||:::   :::         :::|||   ||||||||||   :::      :::
Db        196 GGTAGCTATGATTCTTTGCTTTCAGGTCAAGAATACAATGAT---TTGTCTGCACAGCAT 252

Qy        348 AsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGly 367
              |||   ||||||   |||                    |||   |||:::      |||
Db        253 AATCGGAGAAGGCTATCT--------------------CCTGTTCCTTCTGCTCAAGGT 291

Qy        368 IleLeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrp 387
                  |||||||||   |||  :::                              ||||||
Db        292 TTTAGGCCTTCACCGAGACAAGCATCT----------------------GGTTCATGG 327

Qy        388 GluGlyTyrAspProAlaGlnLeuValArgGluSerLysArgThrArgArgAspGlySer 407
              :::   ||||||      |||      |||::::::||||||:::|||   :::   :::
Db        328 GATGTTTATGACGTGAACCAATTTCAGAGAGATGCAAAACGTTCGAGGATAGAGGCTGCC 387

Qy        408 ------ValAspGlyPheThrProMet-----------------GlyValAspGluArg 419
                    ||||||          |||:::                 |||
Db        388 ATGCCTGTAGATGATCTTGCACCAGTAATTGGTGGGGGTGCTTCAGGTCCATTTGTGAAT 447

Qy        420 SerPheGlyArgGlySerVal----------------------AlaAlaArgProIle 431
                  |||:::|||   |||                              |||      |||
Db        448 GTCCAGGGGAAGGGTCGTGTAAGTCCAACTCCTGGAATGGCCACTGCTGGTGGACCTGGA 507

Qy        432 ArgGlyProProAspSerAspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrPro 451
              |||||||||||||||:::||||||||||:::|||:::|||||||||||||||
Db        508 CTGTCCCCTCCTGATAGTGATTACATATGGAGAGGCATTATTTCTAAGGGTGGAACGCCT 567

Qy        452 ValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluVal 471
              ||||||   ||||||||||||||:::|||||||:::||||||:::::|||||||||
Db        568 GTTTGTCATGCTCGATGTGTTCCCATCGGAAAGGGGTTGGAGACAGAAATTCCTGAAGTT 627

Qy        472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIle 491
              ||||||||||||||||||||   |||:::||||||||||||||||      ||||||
Db        628 GTTAACTGTTCAGCCAGAACAGGATTGGATATGCTTGCAAAACACTATCGTGAGGCTATT 687

Qy        492 GlyCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGlu 511
              |||   :::|||||||||||||:::||||||   |||:::|||||||||||||||||||||
Db        688 GGTTTCGATATTGTTTTCTTCTTACCAGACAGTGAAGACGATTTTGCTTCGTACACTGAA 747

Qy        512 PheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThr 531
              ||||||||||||||   ||||||  ||||||||||||||||||   ||||||||||||
Db        748 TTCCTTCGCTACCTTGGTTCAAAAAC-AGAGCTGGTGTTGCTAAGTTCGATGATGGGACT 806

Qy        532 ThrLeuPhe 534
              |||||||||
Db        807 ACTTTATTT 815

RESULT 13
CA766318
LOCUS       CA766318                800 bp    mRNA    linear   EST 08-JAN-2003
DEFINITION  AF53-Rpf_11_A04_T7_002.ab1 IRRI Drought Stress Panicle Library
            Oryza sativa (indica cultivar-group) cDNA clone C0003844 5' similar
            to unknown, mRNA sequence.
ACCESSION   CA766318
VERSION     CA766318.2  GI:27548346
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Bennett,J., Arumugam,K., Lafitte,R., Wen,J., Rudd,S. and
            Bruskiewich,R.M.
  TITLE     IRRI Drought Stress Panicle cDNA Library
  JOURNAL   Unpublished (2002)
COMMENT     On Dec 2, 2002 this sequence version replaced gi:25995573.
            Contact: Richard Bruskiewich
            Biometrics and Bioinformatics Unit
            International Rice Research Institute
            DAPO 7777, Metro Manila, Philippines
            Tel: +63-2-845-0563
            Fax: +63-2-845-0606
            Email: r.bruskiewich@cgiar.org
            International Rice Information System (IRIS;
            http://www.iris.irri.org): D0203843
            Assignment of putative function to the sequence by S. Rudd of the
            Munich Information Center for Protein Sequences
            (http://mips.gsf.de)
            Plate: 11  row: A  column: 04.
FEATURES             Location/Qualifiers
     source          1. .800
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="IR64"
                     /db_xref="taxon:39946"
                     /clone="C0003844"
                     /tissue_type="Panicles"
                     /dev_stage="Flowering"
                     /clone_lib="IRRI Drought Stress Panicle Library"
                     /note="Vector: pBluescript II SK+; Water stress was
                     applied by not watering for 4 consecutive days. Panicles
                     were collected from control (well watered) and stressed
```

plants at 2 days before heading, at heading, 50% flowering and 4 days after 50% flowering."

ORIGIN

Alignment Scores:

Pred. No.:	2.2e-32	Length:	800
Score:	532.50	Matches:	126
Percent Similarity:	55.51%	Conservative:	25
Best Local Similarity:	46.32%	Mismatches:	84
Query Match:	11.09%	Indels:	38
DB:	6	Gaps:	7

US-09-920-705-3 (1-901) x CA766318 (1-800)

```
Qy        363 ProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMetArgSer 382
              |||         |||||||||||||||||||       |||          |||:::   :::
Db          8 CCTGTTTCTGGTGGTATTCTTCCATCACCGGAATCTGGA------AATCCTCTCTTAACT 61

Qy        383 AsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuValArgGluSerLysArgThr 402
                  :::          :::||||||             |||||||:::||||||
Db         62 GGACGATCTGCTCAGAGTACATTCGATCCA------------AGAGAAGCCAAGAGGTTG 109

Qy        403 ArgArgAspGlySerValAsp--------------------------GlyPheThrPro 413
              |||   |||   :::   |||                          |||
Db        110 CGGTTGGATGCTGCAGCTGACCCTTATGATACAAGGGCAGGTAGTGAGGGTCTTTATTCT 169

Qy        414 MetGlyValAspGluArgSerPheGlyArgGlySerValAlaAlaArgPro--------- 430
                |||      :::|||        |||       ::::::   |||
Db        170 GCTGGGTACTCTCAGCGCGAGAGTGCACGTTCTGAGAGGAGCTCAAGTCCTGCTATTCGA 229

Qy        431 IleArgGlyProProAspSer------------AspHisIleTrpArgGlyMetIleAla 446
              |||   |||                      :::|||   |||||||||   ||||||
Db        230 ATCCATGGCACAGTGCACCGAACATCATACCTTGAGCACTTCTGGCGTGGTAGCATTGCC 289

Qy        447 LysGlyGlyThrProValCysCysAlaArgCysValProMetGlyLysGlyIleGluThr 466
              |||||||||:::|||||||||   |||||||||:::|||:::   ||||||:::|||
Db        290 AAAGGTGGATCTCCTGTATGTCGTGCTCGCTGTTTGCCTATAAGGAAGGGTGTTGAGATA 349

Qy        467 LysLeuProGluValValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHis 486
              ||||||:::|||||||||||||||||||||||||   |||:::|||||||||||||||
Db        350 CCTTTACCGGATGTTGTTAATTGTTCCGCTAGAACTGGACTAGATATGCTGGCAAAGCAT 409

Qy        487 TyrAlaValAlaIleGlyCysGluIleValPhePheValProAspArgGluGluAspPhe 506
              |||       |||   |||   :::|||||||||||||:::||||||   |||:::||||||
Db        410 TATAGAGATGCTTCAGGGTTTGATATCGTCTTCTTCTTGCCAGATAGTGAAGATGACTTT 469

Qy        507 AlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLys 526
              ||||||||||||||||||||||||||||   ||||||   ||||||||||||   |||
Db        470 GTTTCTTACACTGAATTTTTGCGCTACTTAGGCTCAAAAAGTCGGGCAGGGGTTGTAAAG 529

Qy        527 LeuAspAspGlyThrThrLeuPheLeuValProProSerAspPheLeuThrAspValLeu 546
              |||   ||||||||||||||||||||||||||||||||||||||||||   :::   |||
Db        530 GNTGATGGGGGGACTACTTTATTTTTGGTCCCACCTTCCGATTTTTTACGAAATGGTTTG 589

Qy        547 GlnValThrArgGlnGluArgLeuTyrGlyValValLeuLys------------------ 560
              |||           |||||||||||||||||||||||
Db        590 CAAGNTGATGGTCCAGAACGTCTTTACGGTGTAGTATTACATATTCCGCAATGNCTGCTG 649

Qy        561 --------------LeuProProProAlaValProValThrAlaSerTyrArgGlnGlu 575
                           ||||||   |||||||||                        |||
Db        650 CTGCTCCTGCTTCTGCTNCCACTTCCTGCTGTACAAAGGNCACAACTAACTGCACCAGAA 709

Qy        576 SerGlnSerAsnProLeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSer 595
              ||||||                :::::          |||:::   |||
Db        710 TCACAACCT------------TTTCTACG-ATGNAAAGGGAAATTCCTTTGCAGAGAAAG 756

Qy        596 HisSerLeuTyrProProArgGluAsnTyrIleArg 607
                 |||:::      |||      |||:::   |||
Db        757 GATAGTATGATTACNCCTAGCAACAATCACCATCGA 792
```

RESULT 14
CK208950
LOCUS CK208950 1105 bp mRNA linear EST 08-DEC-2003

DEFINITION FGAS020677 Triticum aestivum FGAS: Library 5 GATE 7 Triticum aestivum cDNA, mRNA sequence.
ACCESSION CK208950
VERSION CK208950.1 GI:39571340
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 1105)
AUTHORS Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D., Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A., Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D., Penniket,C., Roach,J.L. and Sarhan,F.
TITLE Functional Genomics of Abiotic Stress In Wheat and Canola Crops
JOURNAL Unpublished (2003)
COMMENT Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon, Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas_ests@cs.usask.ca
This sequence is the direct result of the Base calling software Phred (default parameters). It is the raw base calls. To aid in the identification of the high quality insert the software Lucy (default parameters) has been run on this sequence. Lucy identified the region [13,760].
Plate: L5B015 row: N column: 21.
FEATURES Location/Qualifiers
source 1. .1105
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone_lib="Triticum aestivum FGAS: Library 5 GATE 7"
/note="Vector: pCMV.SPORT6; Crown and developmental stages of spike formation in wheat cultivar Norstar. 4 mRNA populations were combined before constructing the library. The first mRNA population is from 1cm crown sections after 30 days of cold acclimation. The second is from 1cm crown sections after 11 days of deacclimation (before deacclimation plants were fully vernalized for 49 days). The third is from different developmental stages of spike formation (5 to 50mm) that still have not emerged from the leaf (dissection required). The last is from different developmental stages of spike and seed formation after having emerged from the leaf (visible). First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI."
ORIGIN

Alignment Scores:

Pred. No.:	3.54e-32	Length:	1105
Score:	532.50	Matches:	152
Percent Similarity:	50.88%	Conservative:	50
Best Local Similarity:	38.29%	Mismatches:	97
Query Match:	11.09%	Indels:	99
DB:	7	Gaps:	18

US-09-920-705-3 (1-901) x CK208950 (1-1105)

```
Qy        439 HisIleTrpArgGlyMetIleAlaLysGlyGlyThrProValCysCysAlaArgCysVal 458
              |||   |||||||||   |||||||||||||||:::|||||||||   |||||||||:::
Db         87 CACTTCTGGCGTGGAAGTATTGCCAAAGGTGGATCTCCTGTTTGTCGTGCTCGCTGCTTG 146

Qy        459 ProMetGlyLysGlyIleGluThrLysLeuProGluValValAsnCysSerAlaArgThr 478
              |||:::   :::|||||||||       ||||||::::::||||||   ||||||||||||
Db        147 CCTATAAGGAGGGGCATTGAGATACCTTTACCGGACATTGTTAATATCTCCGCTAGAACA 206

Qy        479 AspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGluIleValPhePhe 498
```

```
Db        207 GGATTAGATATGCTGGCAAAACACTATGGAGATGCTTCGGGTTTTGAAATTGTCTTCTTC 266

Qy        499 ValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSer 518
Db        267 TTACCAGACAGTGAAGATGATTTTGTGTCTTACACTGAATTTCTGCGGTACTTGGGCTCA 326

Qy        519 LysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProPro 538
Db        327 AAAAGTCGCGCAGGGGTTGTTAAGGTTGATGCAGGAACCACTTTGTTTTTGGTCCCGCCA 386

Qy        539 SerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValVal 558
Db        387 TCGGATTTCCTGACGAATGTTCTGCAAGTTGATGGTCCAGAGCGCCTTTATGGCGTGGTA 446

Qy        559 LeuLysLeuProPro---------ProAlaValProValThrAlaSerTyrArgGlnGlu 575
Db        447 CTGCACATTCCACAGATGTCCAATGCTGCTGTCCAGAGACCTGTACTGACTGGACCAGAG 506

Qy        576 SerGlnSerAsnProLeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSer 595
Db        507 TCACAAGCT-----------TATTATGAT----------------------GGTGGT 530

Qy        596 HisSerLeuTyrProProArgGluAsnTyrIleArgGlyAlaPro---------GluHis 612
Db        531 GATACCATGCTAGCGATGCAAAGAAATTACAATATGGGCTCTGCTAATGGCAACCATCAT 590

Qy        613 LeuThrAlaAlaSerLysProSerValSerGlu---------------------------- 623
Db        591 CAGGATGCTGATTATCGGGGGTCTTTACGTGAGGAAGCAGTTCAGTCAGGTGTATCAAGC 650

Qy        624 ---ProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGlu 642
Db        651 TTCCCCATG---------AGCCAAATAGCAGGACAGCAAGGACAGTCACTCAAGCCTGAT 701

Qy        643 LeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSer 662
Db        702 ATTATGGCCACATTAGCCAAACTTATGCCA---AATGTGCAGTCATCAGTTCCAGCAAAT 758

Qy        663 HisGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeu 682
Db        759 AGTCAG----------------------GTGGGTAATCTTCAGCAATCCCGTCAGCAG 794

Qy        683 TyrAsnGlyGluAlaProSerGln-----------------AlaTrpLysArgGlyPro 696
Db        795 TTCGGCAGGCAAGCTCCCACCAGCTATGGGGGCATGGTTGGTGCTCAGGAGCAGCACCCA 854

Qy        697 GlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTyrGlyAsnGlnTyrThrPro 716
Db        855 ------ATACATACTGCT--------------------TACAAACCTGAGGTTACCTTA 887

Qy        717 -AlaGlyGlnLeuProProProProSerArgTyrProProAlaSerAsnAsnProAsnTy 736
Db        888 AGCT------TTGCCTCCACAACCT---------CCTCTCCCTTAAAAAGGACCCNTGTT 932

Qy        736 rThrSerGlyMetValHisGlyAsnMetGlnTyrGlnSer------------------- 749
Db        933 T--------------------TCTCTCCCTTACCAGCAGCGATTTTAACCAGAAAACTT 971

Qy        750 ----------GlnSerValAsnMetProGlnLeuSerProLeuProAsnMetProHisAs 766
Db        972 TTACCCCCCTCAAAAAATTAACAGTCCCTTAGTAACAG--TTACCCAGCTTA-------- 1021

Qy        766 nAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProValSerGlnProMetValGl 786
Db       1022 ------------------------------------ATTGCCCGTCCANAAATACA 1041

Qy        786 nGlnTyrGlnProGluAlaSerMetPro---AsnGlnAsnTyrGlyPro 801
Db       1042 ACCTTACCGGCCC---------CTCCCCACCAAAAAGAATGATGGCCCC 1081

RESULT 15
CF322873
```

```
LOCUS       CF322873                582 bp    mRNA    linear   EST 18-AUG-2003
DEFINITION  HDN--02-E23.g1 OsHDAC1-overexpressing transgenic rice lambda phage
            cDNA library II (HDN) Oryza sativa (japonica cultivar-group) cDNA
            clone HDN--02-E23, mRNA sequence.
ACCESSION   CF322873
VERSION     CF322873.1  GI:33793966
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Kim,J.S., Jun,K.M., Cheong,P.J., Kim,M.J., Lee,T.H., Shin,Y.C.,
            Song,S.I., Kim,J.K., Kim,Y.-K. and Nahm,B.H.
  TITLE     Large-scale Sequencing Analysis of Rice ESTs
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Nahm B.H.
            Genomics and Genetics Institute, GreenGene Biotech Inc.; Division
            of Bioscience and Bioinformatics, MyongJi University
            Yongin, Kyeonggi, Korea
            Tel: 82 31 330 6193
            Fax: 82 31 321 6355
            Email: bhnahm@ggbio.com, bhnahm@bio.myongji.ac.kr.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nackdong"
                     /db_xref="taxon:39947"
                     /clone="HDN--02-E23"
                     /tissue_type="callus"
                     /dev_stage="proliferated callus on 2N6 media for 2 weeks"
                     /lab_host="E.coli SOLR"
                     /clone_lib="OsHDAC1-overexpressing transgenic rice lambda
                     phage cDNA library II (HDN)"
                     /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                     XhoI; cDNA was inserted into lambda Uni-ZAP XR vector at
                     5' end with EcoRI and 3' end with XhoI site. mRNA was
                     derived from rice Histone Deacetylase overexpression
                     line."
ORIGIN

Alignment Scores:
Pred. No.:              6.71e-31        Length:         582
Score:                  511.50          Matches:        107
Percent Similarity:     68.21%          Conservative:   26
Best Local Similarity:  54.87%          Mismatches:     49
Query Match:            10.65%          Indels:         13
DB:                     7               Gaps:           5

US-09-920-705-3 (1-901) x CF322873 (1-582)

Qy        438 AspHisIleTrpArgGlyMetIleAlaLysGlyGlyThrProValCysCysAlaArgCys 457
Db          8 GAGCACTTCTGGCGTGGTAGCATTGCCAAAGGTGGATCTCCTGTATGTCGTGCTCGCTGT 67

Qy        458 ValProMetGlyLysGlyIleGluThrLysLeuProGluValValAsnCysSerAlaArg 477
Db         68 TTGCCTATAAGGAAGGGTGTTGAGATACCTTTACCGGATGTTGTTAATTGTTCCGCTAGA 127

Qy        478 ThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGluIleValPhe 497
Db        128 ACTGGACTAGATATGCTGGCAAAGCATTATAGAGACGCTTCAGGGTTTGATATCGTCTTC 187

Qy        498 PheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSer 517
Db        188 TTCTTGCCAGATAGTGAAGATGACTTTGTTTCTTACACTGAATTTTTGCGCTACTTAGGC 247

Qy        518 SerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValPro 537
Db        248 TCAAAAAGTCGGGCAGGGGTTGTAAAGGTTGATGGGGGGACTACTTTATTTTTGGTCCCA 307
```

```
Qy        538 ProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyVal 557
              ||||||||||||||||   :::||||||||||||          ||||||||||||||||||
Db        308 CCTTCCGATTTTTTACGAAATGTTTTGCAAGTTGATGGTCCAGAACGTCTTTACGGTGTA 367

Qy        558 ValLeuLysLeuPro-----------------ProProAlaValProValThrAlaSer 571
              ||||||   :::|||                 |||   :::   |||      |||
Db        368 GTATTACACATTCCGCAAATGTCTGCTGCTGCTCCTGCTTCTGCTCCCACTCCTGCTGTA 427

Qy        572 TyrArgGlnGlu---------SerGlnSerAsnProLeuHisTyrMetAspGlnAlaArg 588
                ||| :::           :::|||  |||      |||  |||:::  |||
Db        428 CAAAGGCCACAACTAACTGCACCAGAATCACAACCT------TACTACGATGAA---AGG 478

Qy        589 AspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGluAsnTyrIleArgGly 608
              :::  |||           :::|||:::      |||      |||:::  |||
Db        479 GAAATTCCTTTGCAGAGAAGGTATAGTATGATTACGCCTAGCAACAATCACCATCGAGAT 538

Qy        609 AlaProGluHisLeuThrAlaAlaSerLysProSerValSerGlu 623
              |||  :::|||      :::      :::   |||:::   :::
Db        539 GCT---GATCACCGTGGATCTTTACGTGAGGATTCATTGCATCAA 580
```

Search completed: April 8, 2005, 13:44:55
Job time : 7010 secs